University of Alberta

A Search for MSSM Higgs Bosons using the OPAL Detector at LEP

by

Anthony Augustine Faust

©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Physics

Edmonton, Alberta

Fall 1999

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-46836-4

Canada

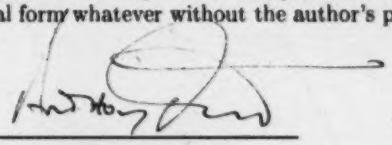
University of Alberta

Library Release Form

NAME OF AUTHOR:	Anthony Augustine Faust
TITLE OF THESIS:	A Search for MSSM Higgs Bosons using the OPAL Detector at LEP
DEGREE:	Doctor of Philosophy
YEAR THIS DEGREE GRANTED:	1999

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Anthony Augustine Faust
11441 University Avenue NW,
Edmonton, Alberta
T6G 1Y9

August 30, 1999

Nothing exists except atoms and empty space;
everything else is opinion.

Democritus of Abdera

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **A Search for MSSM Higgs Bosons using the OPAL Detector at LEP** submitted by **Anthony Augustine Faust** in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

James Pinfold.

Professor James L. Pinfold
(Supervisor)

Robert S. Orr

Professor Robert S. Orr
University of Toronto
(External)

Professor Helmy S. Sheriff
(Chairman)

Assistant Professor Frank A. Hegmann

Assistant Professor Marc de Montigny
Faculté Saint-Jean

Assistant Professor Manuella G. Vincter

DATE: 28th August 1995

Abstract

Using data collected by the OPAL Detector during the 1997 and 1998 LEP runs, a number of Artificial Neural Network analyses were performed to search for Higgs bosons in the Minimal Supersymmetric Model (MSSM) through the production process $e^+e^- \to Z^0 \to h^0A^0 \to A^0A^0A^0 \to 3(b\bar{b})$ at centre of mass energies $\sqrt{s} = 183$ GeV (1997) and $\sqrt{s} = 189$ GeV (1998). No signal was seen and the scalar and pseudo-scalar MSSM Higgs bosons, h^0 and A^0, respectively, were excluded over $2m_{A^0} \leq m_{h^0}$, $m_{h^0} \lesssim 80\,\text{GeV}$, and $m_{A^0} > 12\,\text{GeV}$. The maximal production rate for this process, at 95% CL, admitted by the data was shown to range from $\sigma \times \text{BR}_{(A^0 \to b\bar{b})} \leq 0.192$ pb for $(m_{h^0}, m_{A^0}) = (40, 12)$ GeV to $\sigma \times \text{BR}_{(A^0 \to b\bar{b})} \leq 0.057$ pb for $(m_{h^0}, m_{A^0}) = (80, 40)$ GeV.

Preface

The work presented in this dissertation was carried out at the Centre for Subatomic Research, Department of Physics at the University of Alberta, Edmonton between September 1994 and April 1999, under the supervision of Professor James L. Pinfold.

Results from Sections 3.4 and 3.5 represent original work in the application of Artificial Neural Networks to High Energy Physics data analyses and have been presented to the OPAL community in the form of internal technical reports.

A.A. Faust, *A Search for* $e^+e^- \to h^0A^0 \to A^0A^0A^0 \to 3(b\bar{b})$ *at* $\sqrt{s} = 189\,\text{GeV}$, OPAL-TN/582, Nov., 1998

A.A. Faust and J.L. Pinfold, *A Neural Network Search for* $e^+e^- \to h^0A^0 \to A^0A^0A^0 \to 3(b\bar{b})$ *at* $\sqrt{s} = 183\,\text{GeV}$, OPAL-TN/549, Jun., 1998

Results from Chapters 4 and 5 have been published in the following articles:

OPAL Collab., G. Abbiendi *et al.*, *Search for Neutral Higgs Bosons in* e^+e^- *Collisions at 189 GeV*, Submitted to Euro. Jour. Physics **C**, May 1999.

OPAL Collab., G. Abbiendi *et al.*, *Search for Higgs Bosons in* e^+e^- *Collisions at 183 GeV*, Euro. Jour. Physics **C7** (1999) 407-435.

In addition to these results in the area of data analysis, I contributed to the construction and commissioning of the Time of Flight Endcap (TE) detector in OPAL.

G. Aguillion *et al.*, *Thin scintillating tiles with high light yield for the OPAL endcaps*, Nucl. Instrum. Methods **A417** (1998) 266-277.

I participated in the fabrication and quality control of the optical fibers used in TE and was involved with the data collection during the prototype test beam at CERN. Most recently, I wrote a new version of the high voltage control software for the TE detector.

Acknowledgements

Without the help and patience of many people, this dissertation would not have been possible. A list of only a few of them should not be considered as an underestimation of the contributions of those not named here. For financial support, I would like to gratefully acknowledge The National Science and Engineering Research Council of Canada post-graduate scholarship program and the University of Alberta.

First, I would like to thank my supervisor, Dr. James L. Pinfold, for his interest and enthusiasm both for the research presented here, and also for the many discussions about new and exciting aspects of particle physics. Also to Drs. Bruce Campbell and Alick Macpherson whose patient explanations always left me with a clearer understanding and appreciation of the field.

My dissertation research was carried out as a member of the Higgs Boson working group of the OPAL experiment at LEP. I would like to thank all the members of the working group for their contributions to the development of the common analysis framework which made this analysis possible. Special thanks to Dr. Satoru Yamashita for his interest in my work and to Dr. Wayne Springer for his collaborative effort which led to the results in Section 3.5.

In addition, I would like to thank Norm Buchanan, Andrew Brown, Kipp Cannon, Bryan Caron, Patrick Sutton and the many other students at the University of Alberta who made my time here memorable.

It is the people closest to you that have the greatest impact on your life. They help form your values, instill in you the self-confidence to set your sights high and then give you the encouragement and support to follow them through. My most sincere thanks to my mother, Brenda. You set an example of caring and compassion that I can only hope to emulate. To my grandpa, Douglas and my sister and brother, Catherine and David. Your constant encouragement did not go unnoticed. To my long time friend, Gowan, and his new family. You never hesitated to offer a place to crash or holiday meal when I was away from home. Thanks. I look forward to spending weekends tinkering in an old wooden diesel with a bit of Nelson's blood to keep off the chill.

Lastly, to Jennifer. Your love and unfailing support has changed the direction of my life.

Contents

1 Introduction **1**
1.1 The Standard Model . . . 2
1.1.1 Properties of the Standard Model Higgs Boson . . . 6
1.1.2 Deficiencies of the Standard Model . . . 8
1.2 Beyond the Standard Model . . . 10
1.2.1 The Higgs Mechanism in MSSM . . . 11
1.2.2 MSSM at OPAL . . . 14
1.2.3 The Search for $e^+e^- \to Z^0 \to h^0A^0 \to A^0A^0A^0$ . . . 15

2 LEP and the OPAL Detector **17**
2.1 The Large Electron Positron Collider . . . 17
2.2 The OPAL Detector . . . 18
2.2.1 Central Tracking Chambers . . . 21
2.2.2 Time-of-Flight . . . 33
2.2.3 Calorimetery . . . 37
2.2.4 Muon Detector . . . 41
2.2.5 Forward Detectors . . . 41
2.2.6 The OPAL Trigger and Data Stream . . . 42

3 Artificial Neural Networks **44**
3.1 A Brief History of Neural Nets . . . 45
3.2 The Fundamentals . . . 46
3.2.1 The Neuron . . . 46
3.2.2 The Activation Function . . . 47
3.2.3 Building a Network . . . 48
3.3 Learning with Neural Networks . . . 49
3.3.1 Backpropagation . . . 50
3.3.2 Performance Issues . . . 54
3.4 An Algorithmic Approach to Pruning . . . 58
3.4.1 Pruning of the Input Layer . . . 58
3.4.2 Pruning of the Hidden Layers . . . 60
3.5 Systematic Error Checks with Neural Networks . . . 68
3.5.1 Transformation of Monte Carlo Distributions . . . 68
3.5.2 ANN Systematic Error . . . 73

4 The Search at 183 GeV **74**
4.1 Introduction . . . 74
4.2 Data and Monte Carlo . . . 74
4.3 Preselection . . . 75
4.3.1 Subdetector Status . . . 76

4.3.2 Hadronic Event Selection . . . 76
4.3.3 Event Reconstruction . . . 76
4.3.4 Track-Cluster Matching . . . 78
4.3.5 Jet Reconstruction . . . 79
4.3.6 Topological Variables . . . 80
4.3.7 Initial State Radiation . . . 81
4.3.8 Summary of Preselection Cuts . . . 82
4.4 B Jet Tagging . . . 82
4.4.1 Jet-wise B-Tagging Algorithm . . . 84
4.4.2 Application to the Six Jet Analysis . . . 86
4.4.3 Secondary Vertices . . . 88
4.5 Artificial Neural Network Analysis . . . 90
4.5.1 Topology and Training . . . 90
4.5.2 Training and Testing data sets . . . 90
4.5.3 Network Inputs . . . 91
4.6 Systematic Checks . . . 99
4.6.1 Preselection . . . 99
4.6.2 Si Resolution Smearing and Hit Dropping . . . 99
4.6.3 Different Monte Carlo Generators . . . 103
4.6.4 ANN Systematic Error . . . 107
4.7 Results . . . 111
4.7.1 MSSM Exclusion Limit . . . 121

5 The Search at 189 GeV **122**
5.1 Data and Monte Carlo . . . 122
5.2 Preselection . . . 124
5.2.1 Subdetector Status . . . 124
5.2.2 Event Reconstruction . . . 124
5.2.3 Changes from 183 GeV analysis . . . 124
5.2.4 Summary of Preselection Cuts . . . 125
5.3 B Jet Tagging . . . 130
5.3.1 Application to 6 Jet Analysis . . . 131
5.4 Cut Based Analysis . . . 137
5.5 Neural Network Based Analysis . . . 139
5.5.1 Training and Testing data sets . . . 139
5.5.2 Kinematic Neural Network . . . 141
5.5.3 B-Tagging Neural Network . . . 146
5.5.4 Input Pruning . . . 150
5.5.5 Systematic Checks . . . 150
5.6 Results . . . 156
5.6.1 MSSM Exclusion Limit . . . 171

6 Conclusion **172**

Bibliography **173**

A Derivation of Exclusion Limit **178**

B The OPAL Collaboration **180**

List of Tables

1.1 The fundamental fermions in the Standard Model. 1
1.2 The four known forces and some properties of the exchange bosons. 2
1.3 Gauge charges of the Standard Model fermion fields. 5
1.4 Higgs boson couplings to fermions. 13

3.1 Principal Component Discrimination Analysis applied to the raw inputs of the example distributions 66
3.2 Principal Component Analysis applied to the outputs of the twenty units in the hidden layer 66
3.3 Principal Component Pruning applied to the outputs of the twenty units in the hidden layer. 67

4.1 Monte Carlo and OPAL Data used in the 183 GeV analysis 75
4.2 Detector status values and their interpretation. 76
4.3 Tracks and cluster quality cuts in the preselection 77
4.4 Analysis level event quality selections 78
4.5 Data sets used for the training od neural networks in 183 GeV analysis 91
4.6 Systematic errors due to track parameter smearing 99
4.7 Systematic errors due to Si hit dropping. 99
4.8 Track resolution and Si hit dropping systematic uncertainties for the $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV trained neural network in the 183 GeV analysis 101
4.9 Track resolution and Si hit dropping systematic uncertainties for the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV trained neural network in the 183 GeV analysis 102
4.10 Systematic error in 183 GeV analysis attributed to Monte Carlo generators 104
4.11 Variations in expected background for the two neural networks in the 183 GeV analysis 105
4.12 Results of the Monte Carlo Transformation method applied to the 183 GeV analysis. 110
4.13 183 GeV neural network analysis results 112
4.14 Systematic uncertainties determined for the signal efficiencies and the expected background in the 183 GeV analysis 113

5.1 Data samples used in the $\sqrt{s} = 189$ GeV analysis 123
5.2 The signal efficiencies for the cut based 189 GeV analysis 137
5.3 The predicted Standard Model background for the cut based 189 GeV analysis . . . 138
5.4 Event numbers of those data events selected by the Cut based 189 GeV analysis . . 138
5.5 Data sets used for training the neural networks in the 189 GeV analysis 140
5.6 Principal Component Analysis applied to the outputs of the units in the hidden layer of the Kinematic neural network 142
5.7 Principal Component Pruning applied to the outputs of the units in the hidden layer of the Kinematic neural network 143
5.8 Principal Component Analysis applied to the outputs of the units in the hidden layer of the δ-tagging neural network 146

5.9 Principal Component Pruning applied to the outputs of the units in the hidden layer of the *b*-tagging neural network 147
5.10 Effect of Track Smearing on the signal efficiencies for the neural network based 189 GeV analysis 151
5.11 Effect of Track Smearing on the predicted Standard Model background for the neural network based 189 GeV analysis 151
5.12 Effect of ±400 MeV shift in b-hadron energy spectra on signal efficiencies for the neural network based 189 GeV analysis 153
5.13 Effect of ±400 MeV shift in b-hadron energy spectra on the predicted Standard Model background for the neural network based 189 GeV analysis 153
5.14 Effect of Monte Carlo Transformation on the signal efficiencies for the neural network based 189 GeV analysis 154
5.15 Effect of Monte Carlo Transformation on the predicted Standard Model background for the neural network based 189 GeV analysis 154
5.16 Event numbers of those data events selected by the neural network based 189 GeV analysis 157
5.17 The signal efficiencies for the neural network based 189 GeV analysis 158
5.18 The predicted Standard Model background for the neural network based 189 GeV analysis 158

List of Figures

1.1 Branching ratios of the Standard Model Higgs boson to fermion and gauge boson pairs, including QCD radiative corrections 7
1.2 Higgs production through $e^+e^- \to h^0 Z^0$. Some of the important decay channels are indicated. 7
1.3 Quantum corrections to the Higgs (mass)^2. 9
1.4 The MSSM exclusion for the $\sqrt{s} = 172$ GeV analysis 15
1.5 MSSM Higgs production mechanism in the case $m_{h^0} > 2m_{A^0}$. 16

2.1 Integrated luminosity delivered at LEP2 from 1995-1998. 18
2.2 A cutaway view of the OPAL detector. 20
2.3 Cut-away view of the third phase Silicon Microvertex detector 23
2.4 Impact parameter distributions in the $r - \phi$ plane 25
2.5 Sample event display of an $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ event at LEP2 26
2.6 Cut-away view of the Vertex detector. 26
2.7 Endplate wire layout for a section of the Vertex detector. 27
2.8 Efficiency of resolving a double hit and a double hit associated to a track as a function of the double hit separation in the Jet Chamber 29
2.9 Distribution of $1/p$ for μ^+ and μ^- tracks in $Z^0 \to \mu^+\mu^-$ events collected at the Z^0 peak during the 1998 data run 30
2.10 Intrinsic coordinate resolution, σ_{d_0}, measured using μ^+ and μ^- tracks in $Z^0 \to \mu^+\mu^-$ events collected at the Z^0 peak during the 1998 data run 31
2.11 Measured dE/dx for multihadronic tracks and muon-pairs collected at the Z^0 peak, together with the expected functional form 32
2.12 The Tile Endcap and MIP plug in the OPAL detector 34
2.13 The Tile Endcap sector arrangement. 35
2.14 The MIP plug sector arrangement showing inner and outer layers. 36
2.15 The Tile Endcap detector response to single tracks and the bunchlet identification as seen by the MIP Plug 38
2.16 Electromagnetic calorimeter energy resolution, in GeV, of photons determined from $\ell^+\ell^-\gamma$ events collected at the Z^0 peak 40

3.1 The basic neuron 47
3.2 The general layout of a neural network 49
3.3 The total error as a surface in weight space. 52
3.4 The gradient descent in weight space 53
3.5 Network complexity as related to generalization error 55
3.6 Network complexity as related to training sample size 55
3.7 Evolution of the optimization and generalization errors 57
3.8 Shown are two variables which contain useful discriminatory information for a neural network analysis, but would be missed by a standard cut based analysis 59
3.9 A diagrammatic respresentation of the Pricipal Component Pruning algorithm . . . 65

3.10 The classification efficiencies, between sample signal and background distributions, of a number of networks, differing only in the number of units in the hidden layer . . 69
3.11 Results of Monte Carlo transformation method applied to an example distribution . 72

4.1 A graphic depiction of the hadronization process 79
4.2 Typical dimensions of a B-hadron decay 84
4.3 Schematic presentation of the calculation of the jet-wise b-tagging variable 85
4.4 Comparison between between TSUM and the CLIKE, ULIKE variables 89
4.5 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV 93
4.6 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV 94
4.7 Regulated inputs to the a63 neural network used in the 183 GeV analysis 96
4.8 Regulated inputs to the a83 neural network used in the 183 GeV analysis 97
4.9 Direct comparision of a few neural networks considered in the 183 GeV analysis . . . 98
4.10 Effect of track smearing on Monte Carlo distributions in the 183 GeV analysis . . . 100
4.11 Comparision of PYTHIA and HERWIG predictions for jet resolution parameters in the Durham scheme 106
4.12 Effect of Monte Carlo Transformation technique on distributions of neural network input parameters 108
4.13 Effect of Monte Carlo Transformation technique on distributions of neural network input parameters 109
4.14 The neural network output distributions for the two trained networks in the 183 GeV analysis 114
4.15 Event display of event 12343 in run 8294, selected by the a63 neural network in the $\sqrt{s} = 183$ GeV analysis as a signal candidate. 115
4.16 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV, after application of the a63 network 116
4.17 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV, after application of the a63 network 117
4.18 Event display of event 89200 in run 8712, selected by the a83 neural network in the $\sqrt{s} = 183$ GeV analysis as a signal candidate. 118
4.19 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV, after application of the a83 network 119
4.20 Comparison of neural network input parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 183$ GeV, after application of the a83 network 120
4.21 The MSSM exclusion for the 183 GeV analysis 121

5.1 Distributions of the relevant kinematic analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 126
5.2 Distributions of the relevant kinematic analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 127
5.3 Distributions of the relevant kinematic analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 128
5.4 Distributions of the relevant kinematic analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 129
5.5 B-tagging performance and modelling 131
5.6 Distributions of the relevant b-tagging analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 133
5.7 Distributions of the relevant b-tagging analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 134

5.8 Distributions of the relevant b-tagging analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 135
5.9 Distributions of the relevant b-tagging analysis parameters for OPAL data and the Standard Model prediction at $\sqrt{s} = 189$ GeV 136
5.10 Direct comparision of Kinematic neural networks trained for Principal Component Pruning analysis 144
5.11 Evolution of the Kinematic neural network training 145
5.12 Direct comparision of *b*-tagging neural networks trained for Principal Component Pruning analysis 148
5.13 Evolution of the *b*-tagging neural network training 149
5.14 Sample of analysis variables after the Monte Carlo Transformation technique was applied 155
5.15 Output distributions for the neural networks used in the 189 GeV analysis 159
5.16 Signal efficiencies determined by the neural network based 189 GeV analysis 160
5.17 The signal efficiency surface derived from the neural network based 189 GeV analysis 161
5.18 Event display of event 53614 in run 10311, selected by the 189 GeV neural network analysis as a signal candidate. 162
5.19 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 163
5.20 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 164
5.21 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 165
5.22 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 166
5.23 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 167
5.24 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 168
5.25 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 169
5.26 Distributions of the relevant analysis parameters for OPAL data and the Standard Model prediction, at $\sqrt{s} = 189$ GeV, after the neural network based analysis has been applied 170
5.27 The MSSM exclusion for the 189 GeV analysis 171

6.1 The 2HDM exclusion surface derived from the 189 GeV neural network analysis . . . 172

Chapter 1

Introduction

Early in this century, physicists believed protons, neutrons and electrons were the fundamental building blocks of all matter. Since the Second World War, literally hundreds of particles have been discovered. It was then proposed [1] that protons, neutrons and other previously 'fundamental' particles, were made up of just a few still smaller entities called *quarks*. Electrons, on the other hand, are thought to be fundamental objects themselves and are part of a group of particles known as *leptons*. Over the years, a theory was developed which explained how a small number of quarks and leptons, listed in Table 1.1, collectively known as *fermions*, combine to form all known particles, as well as predicting the existence of a few which have not yet been seen. This theory, known as the Standard Model, also introduces a new class of particles, known as *gauge bosons*, in order to explain the known forces as interactions between these bosons and pairs of the fundamental fermions. Table 1.2 lists the known forces with their associated bosons which are understood to mediate the interactions.

Generation		Particle	Charge	Mass
1	Quarks	up (u)	+2/3	325 MeV
		down (d)	-1/3	325 MeV
	Leptons	electron (e)	-1	511 keV
		ν_e	0	< 17 eV
2	Quarks	charm (c)	+2/3	1.6 GeV
		strange (s)	-1/3	500 MeV
	Leptons	muon (μ)	-1	106 MeV
		ν_μ	0	< 0.27 MeV
3	Quarks	top (t)	+2/3	174 GeV
		bottom (b)	-1/3	5 GeV
	Leptons	tau (τ)	-1	1784 MeV
		ν_τ	0	< 35 MeV

Table 1.1: The fundamental fermions in the Standard Model.

A key feature of the Standard Model is the $SU(3)_{Colour} \times SU(2)_L \times U(1)_Y$ symmetry. Were this symmetry exact, all gauge bosons would be massless, resulting in an infinite range for all the interactions. However, the short range of the Weak force implies its carriers, the Ws and the Z^0, are massive. The presence of the massive exchange bosons indicates symmetry of the original theory has somehow been broken. If the theory is to be renormalizable (that is free of infinities and thus calculable), the breaking of the symmetry must be spontaneous. Within the Standard Model, the spontaneous symmetry breaking is driven by the Higgs mechanism, resulting in a physical scalar, the Higgs boson. Since the Ws and the Z^0 are known from experiment to be massive, the Higgs

Boson	Force	Mass	Range	Spin
photon (γ)	electromagnetism	0	infinite	1
$W^{\pm}$	weak (charged current)	80 GeV	10^{-18} m	1
Z^0	weak (neutral current)	91 GeV	10^{-18} m	1
gluon (g)	strong	0 (assumed)	10^{-15} m	1
graviton	gravity	0 (assumed)	infinite	2

Table 1.2: The four known forces and some properties of the exchange bosons.

mechanism, rather than being a cosmetic detail in the theory, appears to be indispensable. The Higgs boson has, to date, eluded experimental observation.

In this thesis, quantities will be presented in so-called *natural units*, where $\hbar = c = 1$. In this system [length]=[time]=[energy]$^{-1}$=[mass]$^{-1}$. The mass (m) of a particle is therefore equal to its rest energy (mc^2), and to its inverse Compton wavelength ($mc/\hbar$). For example,

$$m_{\text{electron}} = 9.109\times10^{-28}\text{g} = 0.511\,\text{MeV} = (3.862\times10^{-11}\,\text{cm})^{-1}$$

1.1 The Standard Model

The Standard Model (SM), which combines the $SU(2)_L\times U(1)_Y$ Glashow-Weinberg-Salam [2] theory of electroweak interactions, together with Quantum Chromodynamics [3], constitutes a remarkable achievement. The formulation of the theory as a renormalizable quantum field theory preserves its predictive power beyond computations which are first order in the coupling constants (tree-level) and allows for the probing of quantum effects. An array of experimental results have confirmed [4] every feature of the theory to a high degree of precision. In fact, at present there is no compelling evidence which is in conflict with the SM.

The Standard Model of electroweak interactions [5] is based on the gauge group $SU(2)_L\times U(1)_Y$, where the generators of $SU(2)_L$ correspond to the three components of weak isospin T_i and the $U(1)_Y$ generator to the weak hypercharge Y. These are related to the electric charge by $Q = T_3 + Y/2$. The $SU(2)_L\times U(1)_Y$ gauge theory contains three $SU(2)_L$ gauge bosons, W^i_μ, $i = 1,2,3$, and one $U(1)_Y$ gauge boson, B_μ, with kinetic energy terms,

$$\mathcal{L}_{\text{KE}} = -\frac{1}{4}W^i_{\mu\nu}W^{\mu\nu i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\ , \tag{1.1}$$

where

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\nu W^i_\mu - \partial_\mu W^i_\nu + g\epsilon^{ijk}W^j_\mu W^k_\nu\ , \\ B_{\mu\nu} &= \partial_\nu B_\mu - \partial_\mu B_\nu\ . \end{aligned} \tag{1.2}$$

Explicit mass terms for the gauge fields would appear as terms like $\frac{1}{2}m^2_W W_\mu W^\mu$. However, such terms can be shown [6] to violate local gauge invariance.

Masses for the non-abelian gauge fields are generated by the Higgs mechanism [7] via spontaneous symmetry breaking (SSB) which preserves the renormalizability [8] of the theory. The Higgs fields are a complex scalar $SU(2)$ iso-doublet, Φ,

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}\ , \tag{1.3}$$

with a scalar potential given by

$$V(\Phi) = \mu^2\,|\Phi^\dagger\Phi| + \lambda\,|\,\Phi^\dagger\Phi\,|^2\ , \tag{1.4}$$

($\lambda > 0$). This is the most general renormalizable and $SU(2)_L$ invariant potential allowed.

For the choice $\mu^2 < 0$, the ground state of the theory is not at $\Phi = 0$ and the neutral member of the Higgs doublet acquires a vacuum expectation value (VEV) [9]. The direction of the minimum in $SU(2)_L$ space is not determined since the potential depends only on the combination $\Phi^\dagger\Phi$, and we arbitrarily select,

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{1.5}$$

With this choice the scalar doublet has $U(1)_Y$ charge (hypercharge) $Y_\Phi = 1$ and the electromagnetic charge is[1]

$$Q = \frac{(\tau_3 + Y)}{2} . \tag{1.6}$$

Therefore,

$$Q\langle\Phi\rangle = 0 \tag{1.7}$$

so the scalar VEV is neutral and electromagnetism is unbroken, that is, the photon remains massless. The VEV of Equation 1.5 therefore yields the desired symmetry breaking scheme,

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM} . \tag{1.8}$$

It is now straightforward to see how the Higgs mechanism generates masses for the W and Z gauge bosons. The contribution of the scalar doublet to the Lagrangian is,

$$\mathcal{L}_s = (D^\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi) , \tag{1.9}$$

where

$$D_\mu = \partial_\mu + i\frac{g}{2}\tau \cdot W_\mu + i\frac{g'}{2}B_\mu Y \tag{1.10}$$

is the covariant derivative, with g and g' being the $SU(2)_L$ and $U(1)_Y$ coupling strengths, respectively. Expanding Φ about its VEV, the scalar doublet can be written as

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + \mathrm{h}^0 \end{pmatrix} , \tag{1.11}$$

which gives the contribution to the gauge boson masses from the scalar kinetic energy term of Equation 1.9,

$$\frac{1}{2}(0, v)\left(\frac{1}{2}g\tau \cdot W_\mu + \frac{1}{2}g'B_\mu\right)^2\begin{pmatrix} 0 \\ v \end{pmatrix} . \tag{1.12}$$

Rewriting Equation 1.1 in terms of mass eigenstates, the physical gauge fields are then two charged fields, $W^\pm$, and two neutral gauge bosons, Z and γ:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) , \\ Z^\mu &= \frac{-g'B_\mu + gW_\mu^3}{\sqrt{g^2 + g'^2}} , \\ A^\mu &= \frac{gB_\mu + g'W_\mu^3}{\sqrt{g^2 + g'^2}} , \end{aligned} \tag{1.13}$$

with masses from the Higgs mechanism:

$$m_{W^\pm}{}^2 = \frac{1}{4}g^2v^2 , m_{Z^0}{}^2 = \frac{1}{4}(g^2 + g'^2)v^2 , m_A = 0 . \tag{1.14}$$

[1] The τ_i are the Pauli matrices with $Tr(\tau_i\tau_j) = 2\delta_{ij}$.

Since the massless photon must couple with electromagnetic strength, e, the coupling constants define the weak mixing angle θ_W,

$$e = g \sin\theta_W = g' \cos\theta_W \ . \tag{1.15}$$

The measured value of $\sin^2\theta_W = 0.23$ implies that $\tan\theta_W = g'/g \approx 0.5$, so the Weak force is not weak compared to the Electromagnetic force because of its coupling strength. In fact, we see it has a stronger coupling. Rather it is because the weak bosons are massive which causes the Weak force to be short ranged.

It is instructive to count the degrees of freedom after the spontaneous symmetry breaking has occurred. We began with: a complex scalar $SU(2)_L$ doublet Φ with four degrees of freedom; a massless $SU(2)_L$ gauge field, W_i, with six degrees of freedom; and a massless $U(1)_Y$ gauge field, B, with 2 degrees of freedom, for a total of 12. After the spontaneous symmetry breaking there remains a physical real neutral scalar field h^0 (1 degree of freedom); massive W and Z fields (9 degrees of freedom); and a massless photon (2 degrees of freedom). We say that the scalar degrees of freedom have been "eaten" to give the $W^{\pm}$ and Z gauge bosons their longitudinal components.

Fermions can easily be included in the theory. It is convenient to write the fermions in terms of their left- and right-handed projections,

$$\psi_{L,R} = \frac{1}{2}(1 \mp \gamma_5)\psi \ . \tag{1.16}$$

Since the neutrinos are (at least approximately [10]) massless, they can have only one helicity state which is taken to be ν_L. Experimentally, we know that right-handed fields do not interact with the W boson, so the right-handed electron, e_R, must be an $SU(2)_L$ singlet and thus has $Y_{e_R} = -2$. Therefore, to generate the left handed structure of the weak charged current interactions, the $SU(2)_L$ symmetry is applied to left handed quarks (Q) and leptons (L) only. The fermion fields are thus given by,

$$\psi_L: \quad L_{L_k} = \frac{1}{2}(1-\gamma_5)\begin{pmatrix} \nu_k \\ e_k \end{pmatrix}, \qquad Q_{L_k} = \frac{1}{2}(1-\gamma_5)\begin{pmatrix} u_k \\ d_k \end{pmatrix}, \tag{1.17}$$

for the $SU(2)_L$ left-handed doublets and

$$\psi_R: \quad e_{R_k} = \frac{1}{2}(1+\gamma_5)e_k \ , \quad u_{R_k} = \frac{1}{2}(1+\gamma_5)u_k \ , \quad d_{R_k} = \frac{1}{2}(1+\gamma_5)d_k \ , \tag{1.18}$$

for the right-handed singlets, with $k = 1-3$ being a generation index. From Equation 1.6, the hypercharge of the lepton doublet must be $Y_L = -1$. The $SU(2)_L$ and $U(1)_Y$ charge assignments of the first generation of fermions are given in Table 1.3.

Using these hypercharge assignments, the leptons can be coupled in a gauge invariant manner to the $SU(2)_L \times U(1)_Y$ gauge fields,

$$\mathcal{L}_{lepton} = i\overline{e}_{R_k}\gamma^\mu\left(\partial_\mu + i\frac{g'}{2}Y_e B_\mu\right)e_{R_k} + i\overline{L}_{L_k}\gamma^\mu\left(\partial_\mu + i\frac{g}{2}\tau\cdot W_\mu + i\frac{g'}{2}Y_L B_\mu\right)L_{L_l} \ , \tag{1.19}$$

with a similar term for the quark sector.

A fermion mass term takes the form

$$\mathcal{L}_{mass} = -m\overline{\psi}\psi = -m\left(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L\right) \ . \tag{1.20}$$

As is obvious from Table 1.3, the left-and right-handed fermions transform differently under $SU(2)_L$ and $U(1)_Y$ so gauge invariance forbids a term like Equation 1.20. The Higgs boson can be used to give the fermions mass, however. In order to generate a mass term for the fermions we use the fact

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	3	2	$\frac{1}{3}$
u_R	3	1	$\frac{4}{3}$
d_R	3	1	$-\frac{2}{3}$
$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	1	2	-1
e_R	1	1	-2
$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$	1	2	1

Table 1.3: Gauge charges of the Standard Model fermion fields.

that $\Phi^c \equiv -i\tau_2\Phi^*$ is an $SU(2)_L$ doublet, and we can write the $SU(2)_L$ invariant Yukawa coupling of the Higgs boson to the fermions as,

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_{e_k}\bar{L}_{L_k}\Phi e_{R_k} - \lambda_{u_{jk}}\bar{Q}_{L_j}\Phi u_{R_k} - \lambda_{d_{jk}}\bar{Q}_{L_j}\Phi d_{R_k} + h.c.\,, \tag{1.21}$$

Considering the up and down quarks for example, this is,

$$\mathcal{L}_{ud} = -\lambda_d\overline{Q}_L\Phi d_R + h.c.\,, \tag{1.22}$$

which gives the effective coupling

$$-\lambda_d\frac{1}{\sqrt{2}}(\overline{u}_L,\ \overline{d}_L)\begin{pmatrix} 0 \\ v + \mathrm{h}^0 \end{pmatrix} d_R + h.c.\,, \tag{1.23}$$

which can be seen to yield a mass term for the down quark if we make the identification

$$\lambda_d = \frac{m_d\sqrt{2}}{v}. \tag{1.24}$$

For the multi-family case, the Yukawa couplings, λ_d and λ_u, become $N_F \times N_F$ matrices (where N_F is the number of families). Since the fermion mass matrices and Yukawa matrices are proportional, the interactions of the Higgs boson with the fermion mass eigenstates are flavor diagonal and the Higgs boson does not mediate flavor changing interactions. Diagonalization of the masses in the quark sector introduces the weak mixing, or Cabibbo-Kobayashi-Maskawa (CKM) matrix [11]. Similar couplings can be used to generate mass terms for the charged leptons. Since the neutrino has no right handed partner, it remains massless. A complete set of Feynman rules for the interactions of the fermions and gauge bosons of the Standard Model can be found elsewhere [6].

By expressing the fermion kinetic energy, Equation 1.19, in terms of the gauge boson mass eigenstates of Equation 1.13, the charged and neutral weak current interactions of the fermions can be determined. Heralded as the first significant success of the theory, Neutral Current interactions, predicted by the Standard Model as the exchange of a neutral Z^0, were first seen in neutrino scattering experiments in 1972 [12] and the Z^0 boson was directly observed in proton-antiproton collisions in 1983 [13].

One of the most important points about the Higgs mechanism is that all of the couplings of the Higgs boson to fermions and gauge bosons are determined completely in terms of coupling constants and fermion masses. The potential of Equation 1.4 has two free parameters, μ and λ. Rewriting these as,

$$\begin{aligned} v^2 &= -\frac{\mu^2}{2\lambda} \\ m_{\mathrm{h}^0}{}^2 &= 2v^2\lambda\,. \end{aligned} \tag{1.25}$$

The parameter v can be found from the μ decay, $\mu \to e\overline{\nu}_e\nu_\mu$, which proceeds through emission of a virtual W. The strength of this interaction is related to G_F which has been measured very accurately to be $G_F = 1.16639\times10^{-5}\,\mathrm{GeV}^{-2}$. Since the momentum carried by the W boson is of order m_μ it can be neglected in comparison with $m_{W^\pm}$, and we make the identification,

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M_W^2} = \frac{1}{2v^2}, \tag{1.26}$$

which gives the result,

$$v = (\sqrt{2}G_F)^{-1/2} = 246\,\mathrm{GeV}\ . \tag{1.27}$$

With only a single adjustable parameter, m_{h^0}, remaining the Higgs sector of the Standard Model is extremely predictive, with all couplings, decay widths, and production cross sections computed unambiguously in terms of the unknown Higgs boson mass.

Having briefly introduced the essential ingredients of the SM, we now pause to examine its general features and success as a theory. Three principal assumptions went into the building of the theory:

- The gauge group is $SU(3)_C\times SU(2)_L\times U(1)_Y$.
- There is one complex scalar $SU(2)_L$ doublet.
- The fermion representations are left-handed weak isodoublets and right-handed singlets.

In addition to these assumptions the theory contains 21 *a priori* free parameters which must be inserted into the framework of the SM by hand. The missing ingredients of the model are the Higgs boson, which has yet to be discovered, and the τ-neutrino for which there is only indirect evidence at present.

1.1.1 Properties of the Standard Model Higgs Boson

When the mass of the Higgs boson is below the W^+W^- threshold ($\approx 160\,\mathrm{GeV}$), it will decay most often into fermion-antifermion (written $h^0 \to f\bar{f}$) pairs. In the Born approximation, the width into charged lepton pairs is,

$$\Gamma(h^0 \to l^+l^-) = \frac{G_F m_l^2}{4\sqrt{2}\pi} m_{h^0}\beta_l^3\ , \tag{1.28}$$

where $\beta_l \equiv \sqrt{1-4m_l^2/m_{h^0}{}^2}$ is the velocity of the final state leptons. The Higgs boson decay into quarks is enhanced by the color factor, $N_c = 3$, and also receives significant QCD corrections,

$$\Gamma(h^0 \to q\bar{q}) = \frac{3G_F m_q^2}{4\sqrt{2}\pi} m_{h^0}\beta_q^3\left(1+\frac{4}{3}\frac{\alpha_s}{\pi}\Delta_h^{QCD}\right)\ , \tag{1.29}$$

where Δ_h^{QCD} is a QCD correction factor [5]. The Higgs boson clearly decays predominantly into the heaviest fermion kinematically allowed.

For $10\ GeV < m_{h^0} < 160\ GeV$, the most important fermion decay mode is $h^0 \to b\bar{b}$. The electroweak radiative corrections to $h^0 \to q\bar{q}$ are not significant and amount to only a few percent correction [14]. These can be neglected in comparison with the much larger $\mathcal{O}(\alpha_s)$ QCD corrections which, for a Higgs boson in the $100\ GeV$ range, decrease the decay width for $h^0 \to b\bar{b}$ by a factor of about two. The branching ratios for the dominant decays to $f\bar{f}$ pairs are shown in Figure 1.1 [5]. The decrease in the $h^0 \to f\bar{f}$ branching ratios at $m_{h^0} \sim 150\ GeV$ is due to the turn-on of the WW^* decay channel, where W^* denotes a virtual W. For most of the region below the W^+W^- threshold, the Higgs decays almost entirely to $b\bar{b}$ pairs, although decays to $\tau^+\tau^-$ are also used in the experimental searches. The other fermionic Higgs boson decay channels are almost certainly too small to be separated from the backgrounds.

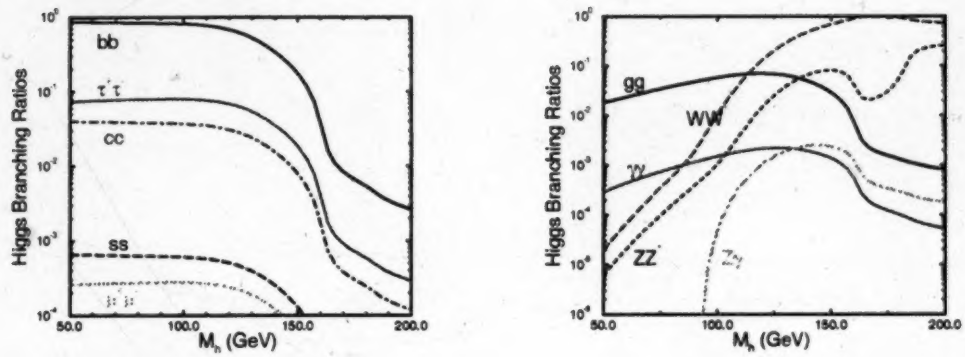


Figure 1.1: Branching ratios of the Standard Model Higgs boson to $f\bar{f}$ (left) and gauge boson (right) pairs, including QCD radiative corrections. The peak in the W^+W^- branching ratio (supressed due to the logarithmic scale) can be seen to correspond to the sharp drop in the $f\bar{f}$ branching ratios.

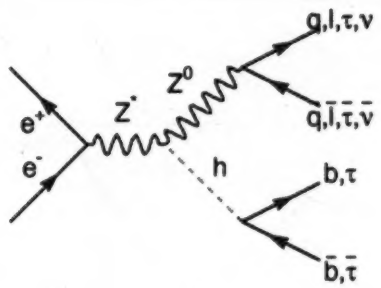


Figure 1.2: Higgs production through $e^+e^- \rightarrow h^0Z^0$. Some of the important decay channels are indicated.

Since the Higgs boson coupling to the electron is very small, $\sim m_e/v$, the $s-$channel production mechanism, $e^+e^- \rightarrow h^0$, is minute and the dominant Higgs boson production mechanism at LEP and LEP2 is the associated production with a Z^0 , $e^+e^- \rightarrow Z^* \rightarrow Z^0h^0$, as shown in Figure 1.2. At LEP2, a physical Z^0 boson can be produced and the cross section is [15],

$$\sigma(e^+e^- \rightarrow Z^0h^0) = \frac{\pi\alpha^2\lambda_{Z^0h^0}^{1/2}[\lambda_{Z^0h^0} + 12\frac{m_{Z^0}{}^2}{s}][1 + (1 - 4\sin^2\theta_W)^2]}{192s\sin^4\theta_W\cos^4\theta_W(1 - m_{Z^0}{}^2/s)^2} , \tag{1.30}$$

where

$$\lambda_{Z^0h^0} \equiv (1 - \frac{m_{h^0}{}^2 + m_{Z^0}{}^2}{s})^2 - 4\frac{m_{h^0}{}^2 + m_{Z^0}{}^2}{s^2} . \tag{1.31}$$

The centre of mass momentum of the produced Z^0 is $\lambda_{Z^0h^0}^{1/2}\sqrt{s}/2$, where $\sqrt{s}$ is the centre of mass interaction energy which, at LEP, is twice the beam energy. The cross section increases rapidly with energy at production threshold and peaks at $\sqrt{s} \sim m_{Z^0} + 2m_{h^0}$, so the best limits on the Higgs boson mass will be obtained at the highest energy.

As mentioned above, a Higgs boson which can be produced at LEP or LEP2 ($m_{h^0} < 110\,\text{GeV}$) will usually decay to $b\bar{b}$ pairs, so the final state from $e^+e^- \to Z^0h^0$ will have four fermions. The dominant backgrounds are $Z^0 \to b\bar{b}(g)$ and $W^+W^- \to q\bar{q}q\bar{q}$ production, which can be efficiently eliminated by b-tagging. LEP2 studies [16] estimate that with $\sqrt{s} = 200\ GeV$ and $\mathcal{L} = 100\ pb^{-1}$ per experiment, a Higgs boson mass of 107 GeV could be observed at the 5σ level.

Currently, the highest energy data at which the four LEP experiments have published a combined Higgs mass limit is at $\sqrt{s} = 183\ GeV$, which is [17],

$$m_{h^0} > 89.8\ GeV \qquad \text{at } 95\%\ \text{CL} \tag{1.32}$$

This limit includes both hadronic and leptonic decay modes of the Z^0.

The successes of the SM as a theory can be listed as:

- Unification of the electromagnetic and weak forces in a manner which maintains the renormalizability and unitarity of the theory.
- Prediction of a specific relationship between W and Z boson masses.
- The weak charged and neutral current structure agrees with experiment.
- All aspects have impressive agreement with all experimental data [4].

1.1.2 Deficiencies of the Standard Model

The SM provides a remarkably successful description of presently known phenomena. The experimental high-energy frontier has advanced into the hundreds of GeV range with no confirmed deviations from Standard Model predictions and few unambiguous hints of additional structure. Still, it seems quite clear that the Standard Model is a work in progress as there remain a number of important questions which the SM does not address. These include:

- The fermion masses and mixings are not predicted.
- Are there massive neutrinos?
- Why are the particles arranged in generations and what sets the number of these generations?
- Charge quantization, i.e. why does $q_e = -q_p$?
- What are the origins of dark matter?
- Why is spacetime 4 dimensional?
- What is the electroweak symmetry breaking mechanism (is there really a light Higgs doublet)?

This list of unanswered questions provides the principle motivation for consideration of physics beyond the SM. Numerous theories are studied in the hope that they will address at least one of these issues. However, no single theory has been invented that successfully addresses all the outstanding questions simultaneously. At the same time, there is not one shred of experimental evidence to provide guidance for extending the theory, or to indicate that physics beyond the SM exists.

Certainly a new framework will be required at the Planck scale $m_P = G_{\text{Newton}}^{-1/2} = 1.2{\times}10^{19}\,\text{GeV}$, where quantum gravitational effects become important. It seems nearly as obvious that new physics exists in the 17 orders of magnitude in energy between the presently explored territory and the Planck scale. The mere fact that the ratio m_P/m_W is so huge is already a powerful clue to the character of physics beyond the SM, this is the infamous *gauge hierarchy problem* [18]. This is not really a difficulty with the Standard Model itself, but rather a disturbing sensitivity of the Higgs potential to new physics in almost any imaginable extension of the Standard Model.

The Standard Model requires a non-vanishing vacuum expectation value (VEV) for the Higgs field, Φ, at the minimum of the potential. Expanding Φ about its VEV yields a single neutral scalar with a classical potential given by

$$V = \mu^2|h^0|^2 + \lambda|h^0|^4 \ , \tag{1.33}$$

where the mass of the physical scalar is $m_{h^0}{}^2 = -\mu^2$. SSB will occur if $\mu^2 < 0$, resulting in $\langle h^0 \rangle = \frac{v}{\sqrt{2}} = \sqrt{m_{h^0}{}^2/\lambda}$. Since we know experimentally that $v = 256\,\text{GeV}$ it must be that $m_{h^0}{}^2$ is very roughly of order $(100\ \text{GeV})^2$. However, $m_{h^0}{}^2$ receives enormous quantum corrections from the virtual effects of every particle which couples, directly or indirectly, to the Higgs field.

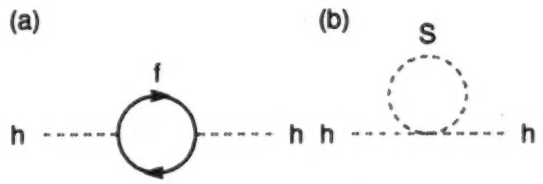


Figure 1.3: Quantum corrections to the Higgs (mass)^2.

For example, in Figure 1.3a we have a correction to $m_{h^0}{}^2$ from a loop containing a Dirac fermion f with mass m_f. The Higgs field couples to f with a term in the Lagrangian $-\lambda_f h^0 f\bar{f}$, thus the Feynman diagram in Figure 1.3a yields a correction [19],

$$\Delta m_{h^0}{}^2 = \frac{\lambda_f^2}{16\pi^2}\left[-2\Lambda^2 + 6m_f^2 \ln(\Lambda/m_f) + \ldots\right] \ . \tag{1.34}$$

Here Λ is an ultraviolet momentum cutoff used to regulate the loop integrals which appear in the evaluation of Figure 1.3a; it should be interpreted as the energy scale at which new physics enters to alter the high-energy behavior of the theory, presumably on the order of the Planck scale or a grand unified scale. Each of the leptons and quarks of the Standard Model can play the role of f; for quarks, Equation 1.34 should be multiplied by 3 to account for color. The largest correction comes when f is the top quark with $\lambda_t \approx 1$.

The problem with Equation 1.34 is that the Higgs boson mass diverges *quadratically*, and this divergence is independent of m_{h^0}. If Λ is of order m_P, say, then this quantum correction to $m_{h^0}{}^2$ is some 30 orders of magnitude larger than the desired value of $m_{h^0}{}^2 \sim (100\ \text{GeV})^2$. There is no symmetry to protect the Higgs mass from these large corrections and, in fact, the Higgs mass wants to be as close to the largest mass scale in the theory. This is only a problem for corrections to $m_{h^0}{}^2$ because quantum corrections to fermion and gauge boson masses do not have the quadratic sensitivity to Λ found in Equation 1.34. However, the quarks and leptons and the electroweak gauge bosons Z^0, $W^\pm$ of the SM all owe their masses to $\langle h^0 \rangle$, so that the entire mass spectrum of the Standard Model is directly or indirectly sensitive to the cutoff Λ.

One could imagine that the solution is to simply pick an ultraviolet cutoff Λ which is not too large. However, one would still have to invent some new physics at the scale Λ which cuts off the loop integral. Furthermore, there is a contribution similar to Equation 1.34 from the virtual effects of any arbitrarily heavy particles which might exist. For example, suppose there exists a heavy complex scalar particle S with mass m_S which couples to the Higgs with a lagrangian term $-\lambda_S h^0 h^0 SS$. Then the Feynman diagram in Figure 1.3b gives a correction

$$\Delta m_{h^0}{}^2 = \frac{\lambda_S}{16\pi^2}\left[\Lambda^2 - 2m_S^2 \ln(\Lambda/m_S) + \ldots\right] \ . \tag{1.35}$$

The term proportional to m_S^2 cannot be eliminated without the physically unjustifiable tuning of a counter-term specifically for that purpose. So $m_{h^0}{}^2$ is sensitive to the masses of the ***heaviest*** particles that h^0 couples to; if m_S is very large, its effects on the SM do not decouple, but instead make it very difficult to understand why $m_{h^0}{}^2$ is so small.

If the Higgs boson is a fundamental particle, we have two options: either we must make the assumption that there exist no heavy particles which couple (even indirectly or extremely weakly) to the Higgs scalar field, or some rather striking cancellation is needed between the various contributions to $\Delta m_{h^0}{}^2$.

The systematic cancellation of the dangerous contributions to $\Delta m_{h^0}{}^2$ can only be brought about by the addition of a new symmetry. It is apparent from comparing Equations 1.34 and 1.35 that the new symmetry should relate fermions and bosons, because of the relative minus sign [2] between fermion loop and boson loop contributions to $\Delta m_{h^0}{}^2$. If each of the quarks and leptons of the Standard Model is accompanied by two complex scalars with $\lambda_S = |\lambda_f|^2$, then the Λ^2 contributions of Figures 1.3a and 1.3b will neatly cancel [20]. Notice that the cancellation occurs independent of the masses, m_f and m_S, and of the magnitudes of the couplings λ_f and λ_S.

Conditions for cancelling all such contributions to scalar masses are not only possible, but are actually unavoidable once we merely assume that a symmetry relating fermions and bosons, called a ***supersymmetry*** [19], should exist.

1.2 Supersymmetry: Beyond the Standard Model

Supersymmetry (SUSY) [21, 22, 23] is considered an attractive extension of the SM, since it provides a solution to the gauge hierarchy problem. Supersymmetry is a symmetry which relates particles of differing spin. Once these particles are combined into ***superfields***, which contain fields differing by one-half unit of spin, the supersymmetry then completely specifies the allowed interactions. In a supersymmetric theory, the scalars and fermions in a superfield have the same couplings to gauge bosons, so the cancellation of quadratic divergences needed to regulate the contributions to the scalar masses occur automatically.

The supersymmetric Lagrangian contains scalar and fermion pairs of ***equal mass***, thus the supersymmetry connects particles of different spin, but with all other characteristics the same. It is clear, then, that ***supersymmetry must be a broken symmetry***. For example, there is no scalar particle with the mass and quantum numbers of the electron. In fact, there are no candidate supersymmetric partners for any of the known fermions or bosons.

Supersymmetric theories are easily constructed according to the rules of supersymmetry, but not all such supersymmetric theories reproduce the successes of the SM. The easiest solution, then, is to start with the SM, and make it supersymmetric in such a way as to preserve its predictive power. The resulting theory, known as the Minimal Supersymmetric Model (MSSM) [24], respects the same $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetries as does the Standard Model. Since there are no candidates for supersymmetric partners of the observed particles, we must double the entire particle spectrum, placing the observed particles in superfields with new postulated superpartners, often called ***sparticles***.

The (s)particle spectrum of the MSSM can be found in a number of references [5, 19, 25]; we will concern ourselves only with the Higgs sector.

[2] λ_S must be positive if the scalar potential is to be bounded from below.

1.2.1 The Higgs Mechanism in MSSM

In the standard (non-supersymmetric) model of electroweak interactions, the fermion masses are generated by the Yukawa terms in the Lagrangian

$$\mathcal{L} = -\lambda_d \overline{Q}_L \Phi d_R - \lambda_u \overline{Q}_L \Phi^c u_R + h.c. \,. \tag{1.36}$$

In a supersymmetric theory, however, a term proportional to Φ^c is not allowed, so another scalar doublet must be added in order to give the $\tau_3 = 1$ components of the $SU(2)_L$ fermion doublets mass. The MSSM, therefore, has two Higgs doublets, Φ_1 and Φ_2. These fields couple separately to up-type quarks for the first doublet, and to down-type quarks and charged leptons for the second doublet. Scalar field doublets that couple in this manner may exist more generally in a class of models known as Type II Two Higgs Doublet Model (2HDM) [25].

The most general MSSM scalar potential can be shown to be [25],

$$\begin{aligned} V_H &= \left(|\mu|^2 + m_1^2\right)|\Phi_1|^2 + \left(|\mu|^2 + m_2^2\right)|\Phi_2|^2 - \mu B \epsilon_{ij}\left(\Phi_1^i \Phi_2^j + \text{h.c.}\right) \\ &\quad + \frac{g^2 + g'^2}{8}\left(|\Phi_1|^2 - |\Phi_2|^2\right)^2 + \frac{1}{2}g^2 |\Phi_1^\dagger \Phi_2|^2 \,. \end{aligned} \tag{1.37}$$

This Higgs potential depends on 3 independent combinations of parameters: $|\mu|^2 + m_1^2$; $|\mu|^2 + m_2^2$; and μB, where B is a new mass parameter. This is in contrast to the general 2HDM doublet model where there are 6 arbitrary coupling constants (and a phase) in the potential. It is interesting to note that the quartic couplings in Equation 1.37 are fixed in terms of the gauge couplings so they are not free parameters, contrary to the SM.

Clearly, if $\mu B = 0$, then all the terms in the potential are positive and the minimum of the potential occurs with $V = 0$ and $\langle \Phi_1^0 \rangle = \langle \Phi_2^0 \rangle = 0$. Hence, all 3 parameters must be non-zero in order for the electroweak symmetry to be broken. The symmetry is broken when the neutral components of the Higgs doublets get vacuum expectation values,

$$\langle \Phi_1^0 \rangle \equiv \begin{pmatrix} v_1 \\ 0 \end{pmatrix} \,, \quad \langle \Phi_2^0 \rangle \equiv \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \,. \tag{1.38}$$

By redefining the Higgs fields, we can always choose v_1 and v_2 positive.

In the MSSM, the Higgs mechanism works in the same manner as in the SM. When the electroweak symmetry is broken, the W gauge boson gets a mass which is fixed by v_1 and v_2,

$$m_{W^\pm}{}^2 = \frac{g^2}{2}(v_1^2 + v_2^2) \,. \tag{1.39}$$

Before the symmetry was broken, the 2 complex $SU(2)_L$ Higgs doublets had 8 degrees of freedom. Three of these were absorbed to give the W and Z gauge bosons their masses, leaving 5 physical degrees of freedom. There is now a charged Higgs boson, $H^\pm$, a CP-odd neutral Higgs boson, A^0, and 2 CP-even neutral Higgs bosons, h^0 and H^0. It is a general prediction of supersymmetric models that there will be an expanded sector of physical Higgs bosons.

After fixing $v_1^2 + v_2^2$ such that the W boson gets the correct mass, the Higgs sector is then described by 2 additional parameters. The usual choice is

$$\tan\beta \equiv \frac{v_2}{v_1} \tag{1.40}$$

and m_{A^0}, the mass of the pseudoscalar Higgs boson. Once these two parameters are given, then the masses of the remaining Higgs bosons can be calculated in terms of m_{A^0} and $\tan\beta$. Note that we can chose $0 \leq \beta \leq \frac{\pi}{2}$ since we have chosen $v_1, v_2 > 0$.

In its most general form, the MSSM has more than one hundred parameters. In this thesis we consider a constrained MSSM, with only six free parameters in addition to those of the SM. The model assumes unification of: the scalar-fermion masses, m_0, at the grand unification (GUT) scale; unification of the gaugino masses, which are parametrized using M_2, the SU(2) gaugino mass term at the electroweak scale; and scalar-fermion tri-linear couplings, A, at the electroweak scale. These simplifications have virtually no impact on the MSSM Higgs phenomenology. Other free parameters of the model are the supersymmetric Higgs mass parameter μ, $\tan\beta$, and the mass of the CP-odd neutral Higgs boson, m_{A^0}.

It is straightforward to find the physical Higgs bosons and their masses in terms of the parameters of Equation 1.37; details can be found in Ref. [25]. The neutral Higgs masses are found by diagonalizing the 2×2 Higgs mass matrix and, by convention, h^0 is taken to be the lighter of the neutral scalar Higgs. This introduces the neutral Higgs mixing angle, α, which enters into many of the couplings and can be expressed in terms of m_{A^0} and β:

$$\tan 2\alpha = \frac{\sin 2\beta(m_{A^0}{}^2 + m_{Z^0}{}^2)}{\cos 2\beta(m_{A^0}{}^2 - m_{Z^0}{}^2)} . \tag{1.41}$$

With our conventions, $-\frac{\pi}{2} \leq \alpha \leq 0$.

At tree level, the masses of the neutral Higgs bosons are given by,

$$m^2_{h^0,H^0} = \frac{1}{2}\left\{ m_{h^0}{}^2 + m_{Z^0}{}^2 \mp \left((m_{A^0}{}^2 + m_{Z^0}{}^2)^2 - 4 m_{Z^0}{}^2 m_{A^0}{}^2 \cos^2 2\beta \right)^{1/2} \right\} . \tag{1.42}$$

The pseudoscalar mass is given by,

$$m_{A^0}{}^2 = \frac{2 \mid \mu B \mid}{\sin 2\beta} , \tag{1.43}$$

and the charged scalar mass is,

$$m^2_{H^\pm} = m_{W^\pm}{}^2 + m_{A^0}{}^2 . \tag{1.44}$$

which give important tree-level predictions about the relative masses of the Higgs bosons,

$$\begin{aligned} m_{H^+} &> m_{W^\pm} , \\ m_{H^0} &> m_{Z^0} , \\ m_{h^0} &< m_{A^0} , \\ m_{h^0} &< m_{Z^0} \mid \cos 2\beta \mid , \end{aligned} \tag{1.45}$$

from which we notice the interesting prediction that the lightest neutral Higgs boson, h^0, is lighter than the Z^0 boson. However, loop corrections to the relations of Equation 1.45 are large. In fact the corrections to $m_{h^0}{}^2$ receive contributions from loops with both top quarks and *squarks* (quark super partners). In a model with unbroken supersymmetry, these contributions would cancel. Since the supersymmetry has been broken by splitting the masses of the fermions and their scalar partners, the neutral Higgs boson masses become at one-loop,

$$\begin{aligned} m^2_{h^0,H^0} = \frac{1}{2}\Bigg\{ & m_{A^0}{}^2 + m_{Z^0}{}^2 + \frac{\epsilon_h}{\sin^2\beta} \pm \Bigg[\left((m_{A^0}{}^2 - m_{Z^0}{}^2)\cos 2\beta + \frac{\epsilon_h}{\sin^2\beta} \right)^2 \\ & + \left(m_{A^0}{}^2 + m_{Z^0}{}^2 \right)^2 \sin^2 2\beta \Bigg]^{1/2} \Bigg\} \end{aligned} \tag{1.46}$$

where ϵ_h is the contribution of the one-loop corrections,

$$\epsilon_h \equiv \frac{3 G_F}{\sqrt{2}\pi^2} m_t^4 \ln\left(1 + \frac{\hat{m}^2}{m_t^2}\right) . \tag{1.47}$$

We have assumed that all of the squarks have equal masses, $\tilde{m}$, and have neglected the smaller effects from the mixing parameters, A_i and μ. For $\tan\beta > 1$, the mass eigenvalues increase monotonically with increasing m_{A^0} and give an upper bound to the mass of the lightest Higgs boson,

$$m_{h^0}{}^2 < m_{Z^0}{}^2 \cos^2 2\beta + \epsilon_h \ . \tag{1.48}$$

The corrections from ϵ_h are always positive and increase the mass of the lightest neutral Higgs boson with increasing top quark mass. The shift in $m^2_{h^0}$ is approximately proportional to $m_t^4 \ln(m_{\tilde{t}}{}^2/m_t^2)$, which can be several tens of GeV. While the top quark mass has been measured to be m_t = (175±5) GeV [26], the masses of the scalar top quarks $m_{\tilde{t}}$ depend on the mixing in the $\tilde{t}$ sector which, in turn, depends on several other parameters of the MSSM.

There are many analyses [27] which include a variety of two-loop effects, renormalization group effects, etc., which found that for given values of $\tan\beta$ and the squark masses, there is an upper bound on the lightest neutral Higgs boson mass. For large values of $\tan\beta$ the limit is relatively insensitive to the value of $\tan\beta$ and with a squark mass less than about 1 TeV, the upper limit on the Higgs mass is about 110 GeV if mixing in the top squark sector is negligible ($A_T \sim 0$). For large mixing, this limit is raised to about 130 GeV. Although this bound is beyond the ultimate reach of the LEP collider, a substantial fraction of this mass range is accessible and provides a definitive test of the MSSM.

Another feature of the MSSM is that the fermion-Higgs couplings are no longer strictly proportional to mass. Rather, these coefficients are determined in terms of the fermion masses and the vacuum expectation values of the neutral members of the scalar components of the Higgs doublets. It is convenient to write the couplings for the neutral Higgs bosons to the fermions in terms of the SM Higgs couplings,

$$\mathcal{L} = -\frac{gm_i}{2m_W}\left[C_{ffh}\overline{f}_i f_i h^0 + C_{ffH}\overline{f}_i f_i H^0 + C_{ffA}\overline{f}_i \gamma_5 f_i A^0\right] , \tag{1.49}$$

where C_{ffh} is 1 for a SM Higgs boson. As $C_{ffA} \to 1$, the $A^0\,f\bar{f}$ coupling is equivalent to the SM values, cf. Figure 1.1, so it decays preferentially to $b\bar{b}$ quark pairs. For large $\tan\beta$ this rate can be significantly enhanced.

f	C_{ffh}	C_{ffH}	C_{ffA}
u	$\frac{\cos\alpha}{\sin\beta}$	$\frac{\sin\alpha}{\sin\beta}$	$\cot\beta$
d	$-\frac{\sin\alpha}{\cos\beta}$	$\frac{\cos\alpha}{\cos\beta}$	$\tan\beta$

Table 1.4: Higgs boson couplings to fermions.

The Higgs boson couplings to gauge bosons are fixed by the $SU(2)_L \times U(1)$ gauge invariance. Two of the phenomenologically important couplings are:

$$\begin{aligned} Z^\mu Z^\nu h^0 : &\quad \frac{igm_{Z^0}}{\cos\theta_W}\sin(\beta-\alpha)g^{\mu\nu} \ , \\ Z^\mu h^0(p) A^0(p') : &\quad \frac{g}{2\cos\theta_W}\cos(\beta-\alpha)(p+p')^\mu \ . \end{aligned} \tag{1.50}$$

We see that the couplings of the Higgs bosons to these gauge bosons depend on the same angular factor, $\beta - \alpha$. The pseudoscalar, A^0, has no tree level coupling to pairs of gauge bosons. The couplings of the neutral scalars to vector bosons ($V = W^\pm, Z$) are suppressed relative to those of the SM

$$g^2_{\Phi_1 VV} + g^2_{\Phi_2 VV} = g^2_{\Phi VV}(SM) \ , \tag{1.51}$$

where $g_{\Phi VV}$ is the coupling of the Higgs boson to vector bosons. Lastly, the Higgs bosons may couple to each other. The important coupling is

$$h^0A^0A^0 : \quad \frac{igm_{Z^0}}{2\cos\theta_W}\cos 2\beta\ \sin(\beta+\alpha)\ . \tag{1.52}$$

A complete set of couplings for all the Higgs bosons at tree level can be found elsewhere [25].

At the current e^+e^- centre of mass energies accessed by LEP, the h^0 and A^0 bosons are expected to be produced predominantly via two processes: the *Higgs-strahlung* process $e^+e^- \to h^0\ Z^0$ (as for the SM Higgs) and the *pair production* process $e^+e^- \to h^0\ A^0$. For these two principal processes, the cross-sections σ_{hZ} and σ_{hA} are related at tree-level to the SM cross-sections [25]:

$$e^+e^- \to h^0Z^0 \ : \ \sigma_{hZ} = \sin^2(\beta-\alpha)\ \sigma^{SM}_{hZ}, \tag{1.53}$$

$$e^+e^- \to h^0A^0 \ : \ \sigma_{hA} = \cos^2(\beta-\alpha)\ \bar{\lambda}\ \sigma^{SM}_{\nu\bar{\nu}}, \tag{1.54}$$

where σ^{SM}_{hZ} and $\sigma^{SM}_{\nu\bar{\nu}}$ are the cross-sections for the SM processes $e^+e^- \to h^0_{SM}\ Z^0$ and $e^+e^- \to \nu\bar{\nu}$, and $\bar{\lambda}$ is a kinematic factor, depending on m_{h^0}, m_{A^0} and $\sqrt{s}$, typically having values between 0.5 and 0.7 for the centre of mass energies under consideration. The coefficients $\sin^2(\beta-\alpha)$ and $\cos^2(\beta-\alpha)$ indicate complementarity of the cross-sections for the two processes, a feature which is exploited in deriving bounds for Higgs boson masses and other model parameters.

1.2.2 MSSM at OPAL

Supersymmetric models have a rich particle spectrum in which to search for evidence of the Higgs mechanism. The various production channels such as $h^0, H^0 \to \gamma\gamma$, $h^0 \to \tau^+\tau^-$, $A^0 \to b\bar{b}$, etc., are sensitive to different regions in the $m_{A^0} - \tan\beta$ parameter space. It takes the combination of many decay channels in order to be able to cover the parameter space completely without any holes. In order to make a statement about the (non)existence of an MSSM Higgs signal in the OPAL data, the cross-sections and branching ratios for every possible combination of allowed MSSM parameters must be calculated and used to predict the number of signal events expected for each search channel. These can then be compared to the experimental result and a confidence level analysis [28] applied to exclude regions of the MSSM parameter space.

Interpretation of the MSSM search results by the OPAL Higgs group was performed centrally, chiefly due to the immense computing requirements. However, in order to motivate the analyses presented in this thesis, a brief outline of the method [28] used to derive the MSSM Higgs boson exclusion plots will be presented.

In scanning the MSSM parameter space, values of the six MSSM parameters introduced on page 12 (m_0, M_2, A, μ, $\tan\beta$, and m_{A^0}), were input to the `HZHA` event generator [29, 30] which is supplemented with parts of the `SUSYGEN` [31] program. The `HZHA` program provides the masses and couplings of all Higgs bosons as well as those of the supersymmetric partners. It also calculates the cross-sections for $e^+e^- \to h^0\ Z^0$ and $h^0\ A^0$ at each centre of mass energy, corrected for initial-state radiation. `SUSYGEN` produces scalar-fermion masses on the electroweak scale, starting from the same input parameters.

The MSSM parameters were varied independently in the following ranges:

- m_0: 0 to 1000 GeV. The masses of physical scalar-fermions are obtained by running m_0 from the GUT scale down to the electroweak scale using the relevant renormalization group equations.
- M_2: 0 to 2000 GeV. The U(1) and SU(3) gaugino mass terms, M_1 and M_3, are calculated from M_2 using the ratios of the corresponding coupling constants, $M_1 : M_2 : M_3 = \alpha_1 : \alpha_2 : \alpha_3$.
- A: $-2.5{\cdot}m_0$ to $2.5{\cdot}m_0$. This range is chosen to include all possible scalar-top mixings.
- μ: -1000 to 1000 GeV.

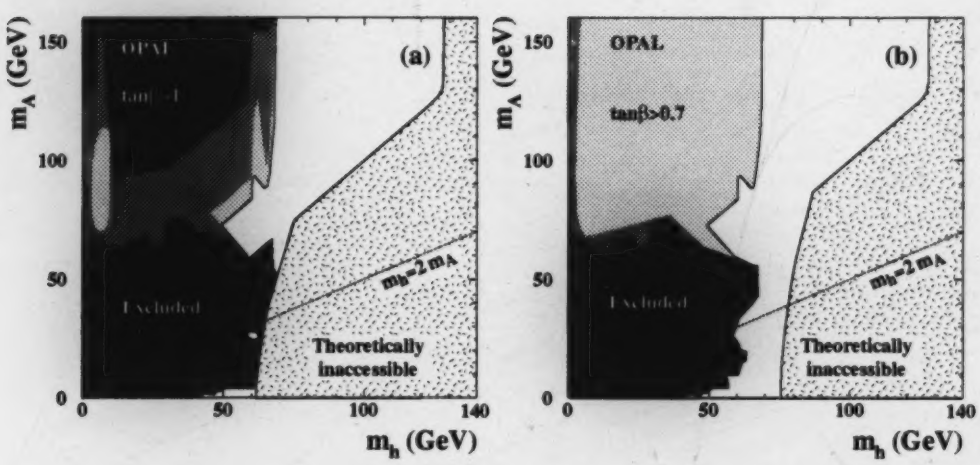


Figure 1.4: The MSSM exclusion for the 172 GeV analysis. Excluded regions are shown for (a) the (m_{h^0} , m_{A^0}) plane for $\tan\beta > 1$, (b) the (m_{h^0} , m_{A^0}) plane for $\tan\beta > 0.7$. All exclusion limits are at 95% CL. The black areas are excluded soley by comparing data to the expected MSSM production rates, without applying any additional theoretical criteria. The speckled areas are theoretically inaccessible.

- m_{A^0}: 5 to 160 GeV. Beyond this range, values in 5 GeV slices around $m_{A^0} = 250$, 400, 1000 and 2000 GeV were also explored.
- $\tan\beta$: 0.7 to 50. This spans the theoretically favoured range, $1 < \tan\beta < m_t/m_b$, but also includes values less than unity, which are not ruled out by theory.
- m_t: 165, 175 and 185 GeV. As mentioned previously, the top quark mass has a strong impact on m_{h^0}: Therefore, it is also varied within reasonable bounds. The range of values includes approximately two standard deviations of the measured top quark mass.

Most parameters are scanned by dividing their ranges into bins of variable size and choosing the values in each bin at random. Exceptions are the parameter m_t, for which three discrete values are used, and the values of m_{A^0} greater than 160 GeV, for which 5 GeV bands around the values are used. The number of parameter sets considered in this scan of the MSSM parameter space was close to 6,000,000.

The results from the search with the OPAL detector for neutral Higgs bosons in the MSSM model at $\sqrt{s} \leq 172$ GeV [28] are reported in Figure 1.4, in the form of experimental exclusion limits, at the 95% CL.

1.2.3 The Search for $e^+e^- \to Z^0 \to h^0A^0 \to A^0A^0A^0$

In Figure 1.4, one can see regions of unexcluded parameter sets with small values of m_{h^0}. These points are characterized by a small value of $\sin^2(\beta - \alpha)$, hence the h^0 Z^0 production is suppressed (cf. Equation 1.53). The h^0 A^0 production is kinematically allowed at LEP2 energies in most of the affected region, but the cross-section is small, requiring a dedicated search to improve tagging efficiency.

The analyses presented in this thesis were conducted in order to close the unexcluded regions satisfying $2m_{A^0} \leq m_{h^0}$. For $2m_{A^0} \leq m_{h^0}$, the process $h^0 \to A^0$ A^0 is also allowed and may even be the dominant decay, leading to more complex final states than those from direct decays into fermion pairs. Since the neutral Higgs bosons in the mass range accessible by LEP2 couple to

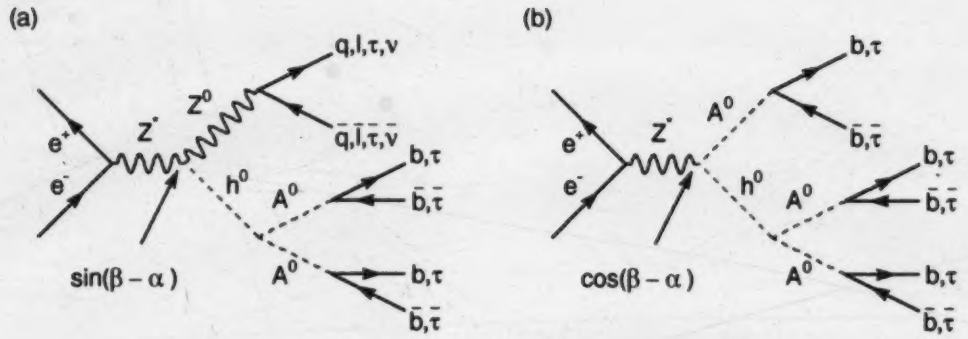


Figure 1.5: MSSM Higgs production mechanism in the case $m_{h^0} > 2m_{A^0}$.

fermions with a strength proportional to the fermion mass, the A^0 dominantly decays into pairs of the most massive fermions which are allowed by the kinematics, most notably into $b\bar{b}$ quark pairs. Figure 1.5 presents the production processes leading to a three Higgs boson intermediate state. As in the case of $h^0 Z^0$ and $h^0 A^0$ production, Equations 1.53 and 1.54, the $h^0 Z^0 \rightarrow A^0 A^0 Z^0$ and $h^0 A^0 \rightarrow A^0 A^0 A^0$ are complementary in nature and so the suppression of $h^0 Z^0$ production, as mentioned above, leads to an increase in the associated $h^0 A^0$ production. We are then searching for the process depicted in Figure 1.5b, where the Higgs bosons decay into $b\bar{b}$ quark pairs. This is an intriguing event topology with no less than six primary jets expected and significant B-hadron production.

It is clear then that the search for $A^0 A^0 A^0 \rightarrow 3(b\bar{b})$ will require good jet resolution with efficient B-hadron identification, both of which depend heavily on the tracking efficiency of the detector. The next chapter will describe the OPAL detector, with emphasis on the central tracking system. This will be followed by a chapter which develops a method of event classification through the use of Artificial Neural Networks. The final two chapters present the results of the search for $e^+e^- \rightarrow h^0 A^0 \rightarrow A^0 A^0 A^0 \rightarrow 3(b\bar{b})$ at $\sqrt{s} = 183\,\text{GeV}$ and $189\,\text{GeV}$, respectively.

Chapter 2

LEP and the OPAL Detector

2.1 The Large Electron Positron Collider

The Large Electron Positron (LEP) collider [32] is located at the European Laboratory for Particle Physics (CERN [1]), near Geneva, Switzerland. LEP accelerates electrons and positrons in a circular underground tunnel 26.7 km long, and collides them at centre of mass energies up to around 200 GeV (196 GeV achieved in 1999). During the experimental run in 1996-1997, LEP operated primarily at 183 GeV, rising to 189 GeV for the 1997-1998 season.

The LEP collider itself is actually the last step in a series of particle accelerators at CERN. Electrons are initially accelerated to 200 MeV by the LEP Injector Linac (LIL). These electrons may be directed onto a target to produce positrons. The electrons (or positrons) are then accelerated to 600 MeV before entering the Electron Positron Accumulator (EPA), where they are stored until eight compressed bunches have been accumulated. The beam is then injected into the Proton Synchrotron (PS), where the energy is increased to 3.5 GeV. Next, the beam energy is increased to 20 GeV by the Super Proton Synchrotron (SPS). Finally, the beam enters the LEP accelerator and collider.

The LEP accelerator [33] consists of eight 2.8 km long arcs linked by eight straight sections. The particles are kept in their orbits in the arcs by 3400 bending magnets and are focussed by 800 quadrupoles and 500 sextupoles. The only difference in this process between electron beams and positron beams is that LEP accelerates the two beams in opposite directions. In the first phase of the LEP programme (LEP1), which began in 1989, electrons and positrons were accelerated to centre of mass energies of approximately 91 GeV (45.5 GeV per beam) using copper radiofrequency (RF) accelerating cavities, located at diametrically opposite positions in the straight sections on either side of the underground experimental halls. Once the beams were at collision energy, these RF cavities then provided energy to compensate for the energy loss due to synchrotron radiation in the arcs. This loss is proportional to E^4_{Beam} and was about 125 MeV per turn at LEP1 energies.

Operation at this energy continued until 1995 when the RF system underwent an upgrade for the second phase of LEP (LEP2) by the installation of superconducting RF cavities, initially increasing the centre of mass energy to about 133 GeV. Further upgrades have resulted in collisions at centre of mass energies of 161 GeV, 172 GeV, 183 GeV, 189 GeV, and currently 196 GeV, which has been attained using the complete installation of 272 superconducting RF cavities operating at average gradients of 7 MV/m to compensate for the 2049 MeV lost per turn due to synchrotron radiation. The maximum LEP luminosity (number of particles per unit area per unit time) that was achieved in 1998 was $8.42 \times 10^{31} \mathrm{cm}^{-2} \mathrm{sec}^{-1}$.

The electron and positron beams are made to collide at four interaction points around LEP. Large, general purpose detectors (OPAL [34, 35], ALEPH [36], DELPHI [37], L3 [38]) were built

[1] Known in French as "Conseil Européen pour la Recherche Nucléaire".

in the large underground halls, 70 m long and 23 m in diameter, at each of these interaction points. Electrostatic separators prevent the beams from colliding at locations where no detectors are situated.

The beams have a lifetime of about 6 hours, during which they are stable enough to do physics. The beams travel in two counter rotating sets of eight bunches per beam, known as 8×8 mode. At the interaction points, these bunches are gaussian in shape with dimensions [39] of $\sigma_x \approx 200$ μm, $\sigma_y \approx 4$ μm and $\sigma_z \approx 15$ mm. The luminosity, $\mathcal{L}$ [cm^{-2}sec^{-1}], of a collider like LEP is given by,

$$\mathcal{L} = \frac{k_b f_o N^+ N^-}{4\pi\sigma_x\sigma_y}\,, \tag{2.1}$$

where k_b is the number of bunches, f_o the bunch rotation frequency ($\approx$ 11 kHz) and $N^{+(-)}$ the number of particles per $e^+(e^-)$ bunch ($\mathcal{O}(3\times10^{11}$/Bunch)).

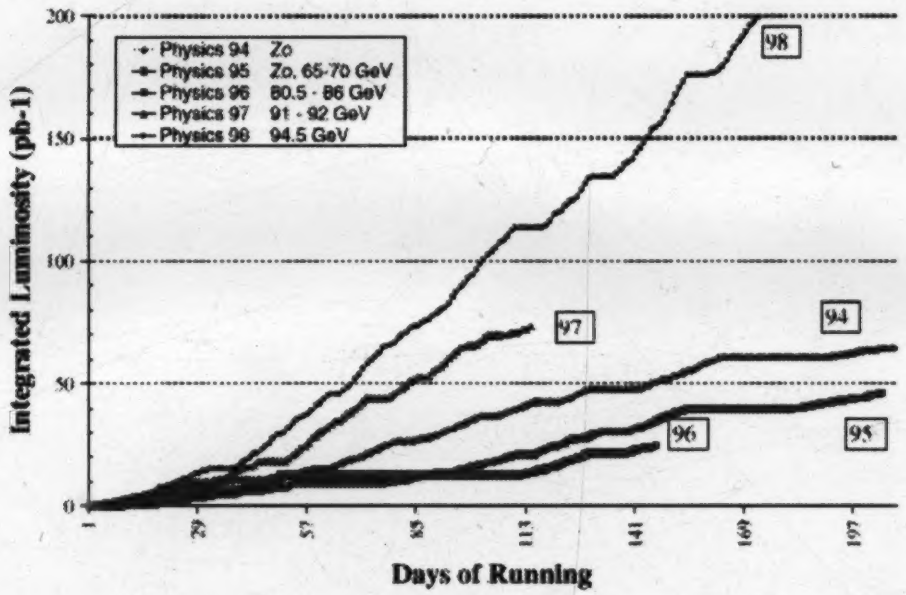


Figure 2.1: Integrated luminosity delivered at LEP2 from 1995-1998.

The rate for a particular reaction is related to the luminosity by $R = \sigma\mathcal{L}$, where σ is the corresponding cross-section. The total number of events collected from this particular reaction, after a period T of collision time, is then $\langle N\rangle = \sigma\mathcal{L}_{\rm int} = \sigma\int_0^T \mathcal{L}{\rm d}t$, where $\mathcal{L}_{\rm int}$ is the integrated luminosity. Henceforth, the luminosity will be understood to mean the integrated luminosity.

2.2 The OPAL Detector

The OPAL (Omni-Purpose Apparatus for LEP) detector is one of four detectors located at the LEP e^+e^- storage ring. It is designed to provide efficient particle detection, with accurate and unambiguous event reconstruction of all possible interactions occurring in e^+e^- collisions. The general layout of the detector is shown in Figure 2.2.

The main features of OPAL are:

- Tracking of the trajectories of charged particles in a uniform magnetic field with measurements of their direction and momentum; particle identification using $\mathrm{d}E/\mathrm{d}x$; and reconstruction of primary and secondary vertices using the central tracking and vertex detectors.
- Measurement of energy and direction of photons and electrons using electromagnetic calorimeters.
- Measurement of hadronic energy by total absorption in the instrumented magnetic return yoke (the hadron calorimeter).
- Identification of muons using muon chambers to measure the position and direction of particles which have passed through the hadron calorimeter.
- Measurement of the absolute machine luminosity using Bhabha scattering events in the forward direction, with respect to the beam line, using the forward detector or the silicon-tungsten luminometer.

The OPAL collaboration (cf. Appendix B) currently consists of over 300 physicists and technicians, from over 30 institutions. The numerous development, construction, maintenance and operation tasks are divided among the participating institutions.

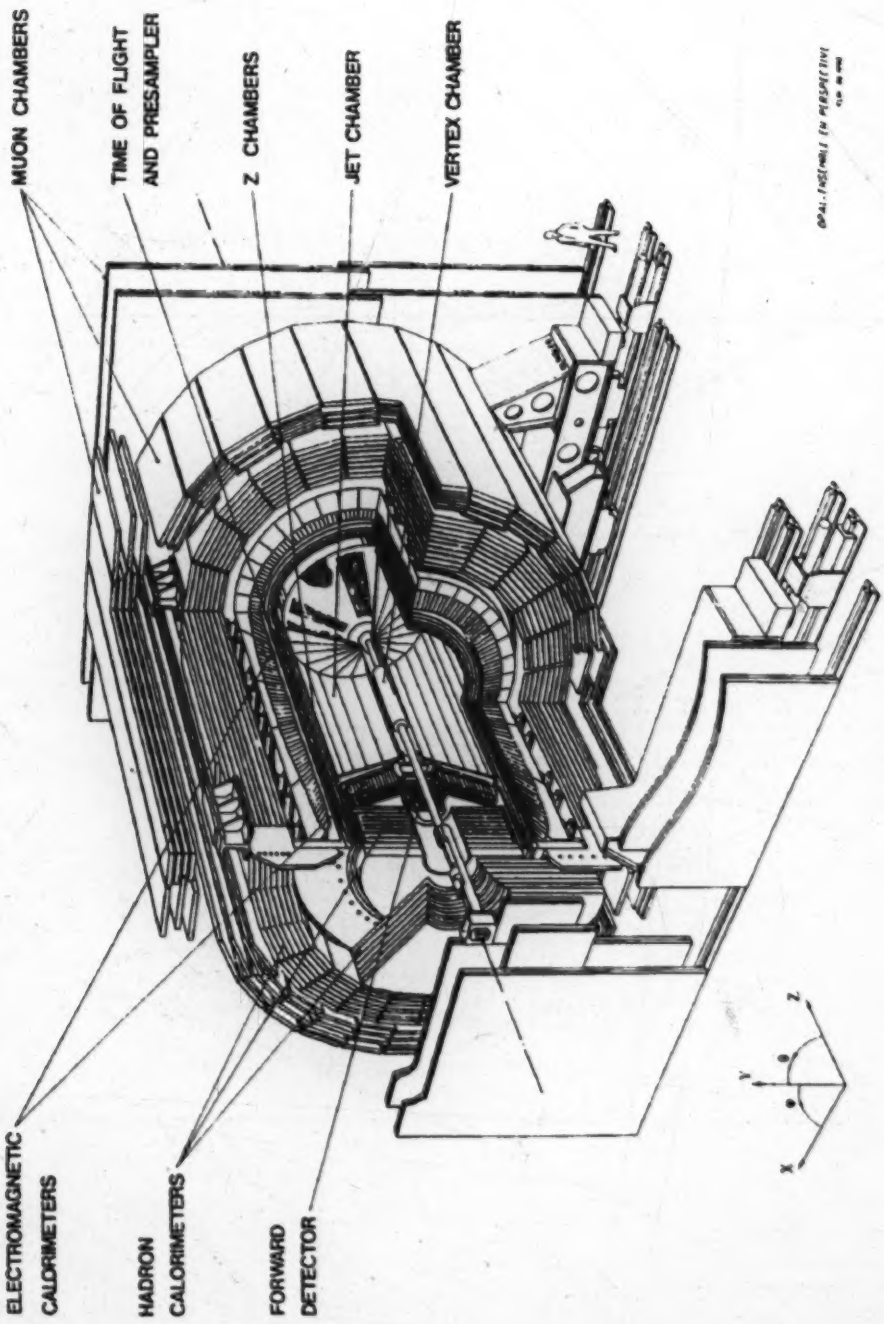


Figure 2.2: A cutaway view of the OPAL detector.

The OPAL Magnet

The magnet consists of a solenoidal coil and iron return yoke. The return yoke is constructed of soft steel plates and can be dismantled into five parts; a central part, two 'C'-shaped parts and two poletips. In the central cylindrical volume, which is used for tracking, the magnetic field is approximately 0.435 Tesla. In the jet chamber tracking volume, which constitutes the bulk of the central tracking system, the magnetic field is measured to be uniform to within $\pm 0.5\%$.

The OPAL Coordinate System

The z-direction is along the beam direction and in OPAL this coincides with the direction of the electrons (which is anti-clockwise when LEP is viewed from above). The x-direction points towards the centre of the LEP ring. The y-direction is normal to the $x-z$ plane, such that the three directions form a right-handed coordinate system. The z-direction is inclined by 1.39° with respect to the horizontal and therefore the y-direction is inclined by 1.39° with respect to the vertical. It is common for a mixture of cylindrical and spherical coordinates to be used, where the z-direction is the same as defined above, the θ-direction is the polar angle with respect to the positive z-axis and ϕ is the azimuthal angle which is 0 rad along the x-direction and $\pi/2$ rad along the y-direction.

The OPAL Track Parameters

The trajectories of charged tracks in the OPAL central detector are specified using a set of parameters ($\kappa, \phi_0, d_0, \tan\lambda$, and z_0), defined as follows:

- κ is the curvature, with $\kappa = 1/(2\rho)$ and ρ the radius of curvature. In OPAL, with the magnetic field along the positive z axis, a positive value of κ corresponds to a particle with negative charge.
- ϕ_0 is the angle made by the track tangent at the point of closest approach (pca) to the origin in the $x-y$ plane.
- d_0 is the track impact parameter. $|d_0|$ is the distance from the origin to the point of closest approach (pca) in the $x-y$ plane. If $\mathbf{d}$ is the vector from the origin to the pca, $\hat{\phi}$ the unit vector at the pca and $\hat{\mathbf{z}}$ the unit vector along the z axis then $d_0 = \hat{\phi} \times \mathbf{d} \cdot \hat{\mathbf{z}}$ gives the sign of the impact parameter.
- $\tan\lambda$ is the *dip angle*, equal to $\cot\theta$ where θ is the polar angle measured from the z axis.
- z_0 is the value of z at the pca. The $s-z$ projection of the track is a straight line, $z = z_0 + s\tan\lambda$, where s is the $r-\phi$ path length from the pca.

With a magnetic field of B kGauss and κ in units of cm^{-1} the momentum, in GeV, of the particle in the $r-\phi$ plane is given by $p_t = 1.5\times10^{-4}B/\kappa$. The three components of the momentum are then given by $p_x = p_t\cos\phi_0$, $p_y = p_t\sin\phi_0$ and $p_z = p_t\tan\lambda$.

2.2.1 Central Tracking Chambers

Of primary importance to the analysis presented in this thesis, the OPAL central tracking system consists of: a silicon microvertex detector (SI); a precision vertex detector (CV); a large volume jet chamber (CJ); and z-chambers (CZ), all contained within the solenoid of the magnet.

The gas used in the central tracking detectors, which is common to the CJ, CV and CZ, is a mixture consisting of 82% argon, 9.8% methane and 2.0% isobutane, contained under a pressure of 4 bar within a pressure vessel whose inner wall is at a radius of 7.8 cm from the interaction point. This in turn houses the beampipe and silicon microvertex detector. The gas pressure is chosen to optimize [40] particle separation using $\mathrm{d}E/\mathrm{d}x$ and is a compromise between high pressure, to

minimize diffusion effects which degrade spatial resolution at long drift distances, and low pressure, to minimize multiple scattering which is conducive to good momentum resolution.

Silicon Microvertex Detector

A critical property of modern detectors is the ability to reconstruct the displaced vertices generated by the decay of particles with typical decay lengths $\mathcal{O}(1\ mm)$. This is best achieved by building high resolution detectors as close to the primary interaction point as possible. This was the motivation for the construction of the vertex detector (Section 2.2.1) which is a high resolution wire chamber sitting immediately outside the beampipe. However, in practice there are practical limits to how close to the interaction point the detector can be instrumented. The chief concerns being the damaging radiation flux, which increases closer to the interaction point, and the desire to minimize the radiation lengths of material particles must traverse before interacting with the calorimetry. For high rate environments, solid state detectors are required, but in the original OPAL design the beampipe was large enough so that background rates were not a limiting issue and the more standard Vertex detector was sufficient.

After the first LEP run in 1989, it was found that the lower than expected background rate permitted the use of a smaller diameter beampipe. With the reduced beampipe a thin annular region was created between the new beampipe, with an outer diameter of 56.5 mm, and the inner wall of the wire chamber pressure vessel, with an inner diametre of 80 mm. This permitted the installation of a new precision solid state detector closer to the interaction region.

The first silicon microvertex detector (SI) [41] was added to OPAL in the 1990/91 winter shutdown, in between the original beampipe and a new 0.11 cm thick beryllium pipe at a radius of 5.35 cm. The first model began taking data in June 1991 and provided increased precision in the ϕ measurement (the r measurement was known from the surveyed position of the detector). In 1993 a second model [42] was installed which used the same technology as the first but returned both z and ϕ measurements.

The excellent single hit resolutions of the first SI detectors were demonstrated in Z^0 di-lepton decays, where the impact parameter resolutions were $\sigma_{d_0} = 15\,\mu$m in the $r-\phi$ plane, and $\sigma_{z_0} = 20-50\,\mu$m in the z plane, for polar angles ranging from $|\cos\theta| = 0.0-0.8$. The additional precision in the measured track parameters significantly improved the measurements of the τ lepton [43] and b flavoured hadron lifetimes [44], as well as the identification of b flavoured multi-hadron events. The additional z information provided by the second model detector allowed for three dimensional track fitting and significantly reduced the number of tracks falsely associated with reconstructed vertices.

In autumn 1995, the LEP collider entered a new era, known as LEP2, in which the centre of mass energy is being raised in steps from the Z^0 resonance to ≈ 200 GeV. One of the main goals of the LEP2 phase of operation is the search for the Higgs boson. The most popular Higgs scenarios predict that the Higgs boson(s) will decay predominately into $b\bar{b}$ quark pairs. To distinguish Higgs bosons from irreducible backgrounds due to Z^0 and W^+W^- decays, b quark tagging is an indispensable tool, from which the highest possible tagging efficiency and purity is desired. In this light, it was felt that the SI detector should be upgraded to provide the largest possible geometrical acceptance. This third phase of the SI detector was installed for the 1995 data run, and it is this latest model which will be presented here. The detector is fully described elsewhere [45].

The basic modular unit of the detector is the *ladder*. The ladder consists of three back-to-back $z\ \phi$ readout silicon wafers aligned lengthwise in a row, as shown in Figure 2.3. Twelve ladders form the inner cylindrical detection layer and fifteen comprise the outer layer, together providing single hit coverage within $|\cos\theta| \leq 0.93$ and double hit coverage over $|\cos\theta| \leq 0.89$.

In azimuthal coverage, the ladders are arranged with a small tilt angle with respect to the circumference so that the detectors have a small overlap. The available radial space does not allow an overlap in the sensitive detector area, but the dead region between the ladders in the ϕ coordinate is minimized. This geometry yields 97% single hit coverage within the double layer

region, $|\cos\theta| \leq 0.89$.

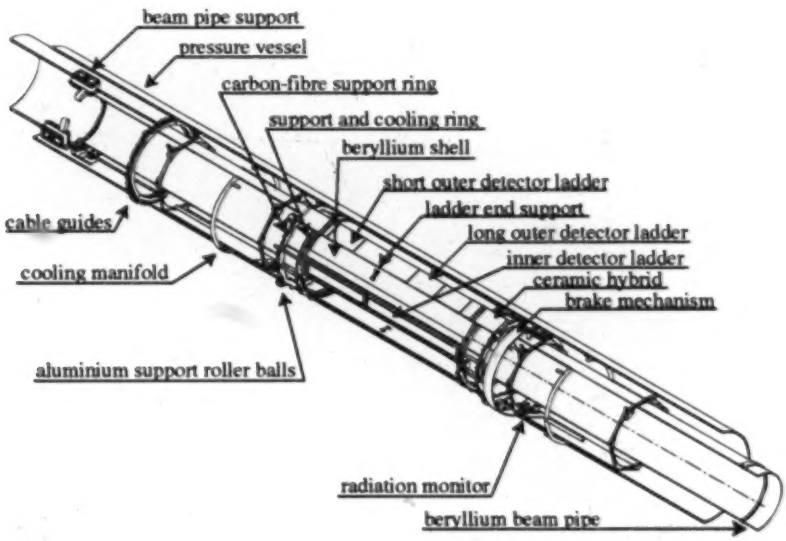


Figure 2.3: Cut-away view of the third phase Silicon Microvertex detector, installed in 1995.

The main components of the ladders are the silicon wafers with strip readout. In the original detector, each wafer supported 629 semi-conductor readout strips. In the upgraded model, each wafer is, in fact, made of two layers. The first layer is a 250 μm thick high resistivity (> 5 $k\Omega$ cm) n-type silicon wafer with a metallization on the backplane. Between the silicon and the backplane is a thin protective layer of n^+ doped silicon. On the front (readout) side are the ϕ readout strips. Adhered to that is a similar layer but with the readout strips running orthogonal to the previous set, thus yielding the third coordinate measurement.

The offline cluster finding is performed separately in ϕ and z. A ϕ cluster is required to have at least one strip (the *seed*) with a pedestal-subtracted pulse height greater than 4σ where σ is the rms noise of the individual strip. Adjacent strips are included if their pulse heights exceed 2σ. Clusters in the z coordinate may include 2 strips on either side (provided their pulse heights are at least 2σ) in allowance for the case where the incident particle traverses the detector thickness at a wide angle.

The ϕ and z impact positions are then determined from the mean of the cluster strip positions weighted by their pulse heights. These impact positions are then converted into hit positions within the OPAL detector and result in one ϕ and three z hits, which is due to the three-fold ambiguity in the z position because of the shared readout of the three z wafers on each ladder. The error assignment is then the quadratic sum of the ϕ and z uncertainties.

One of the main obstacles in designing the z position measurement system was to incorporate the readout without adding significant material within the acceptance of the detector in contrast to putting the z-readout electronics directly at the end of the z strips. Successful results have been achieved via routing the z strip signals, to the front-end electronics at the end of the ladders, through metallized prints on glass. The OPAL design uses a gold printed circuit on a glass substrate. The 200 μm thick glass substrate was chosen to minimize material in the active region. The wafer dimensions are 33×66 mm.

Care was taken to minimize the radiation length of material such that for a particle emerging transverse to the beam line, the total amount of material traversed in detector is, on average, 1.5% of a radiation length at normal incidence. The main contributions are from the silicon detectors (1.1%), the z-readout substrate (0.3%) and the beryllium support shells (0.1%).

The precise knowledge of the position and orientation for the earlier phases of the detector were acquired mostly by a software alignment procedure which utilized $Z^0 \rightarrow \ell^+\ell^-$ decay data. Although the ultimate alignment information still relies on such a procedure, the production rate of tracks needed for software alignment is greatly reduced for data collected at energies above the Z^0 peak. Nevertheless, a fast detector calibration is desirable for rapid analysis of new data at high energies. For these reasons the phase three detector employed a detailed optical survey prior to its installation to ensure a reasonably accurate starting point for the software alignment procedure. The ultimate position uncertainty resulting from the alignment was estimated to be about $8-10\,\mu$m in the $r-\phi$ plane and $10-12\,\mu$m in $r-z$ coordinates and is limited by the systematic errors of the overall OPAL tracking system.

An indication of how the detector improves the overall OPAL tracking performance is given by the apparent separation at the interaction point of two tracks in a dilepton event, from which the impact parameter resolutions σ_{d_o} and σ_{z_o} can be determined. Using the SI information, combined with information from the rest of the tracking system, a resolution [45] of $\sigma_{d_o} = 18\,\mu m$ is achieved for tracks with $|\cos\theta| \leq 0.75$. A resolution of $\sigma_{z_o} = 24\,\mu m$ is achieved for tracks emerging in the region normal to the beam axis (cf. Figure 2.4). These values are slightly worse that those found for the second phase detector [42] which can be attributed to the lower charged track statistics available at LEP2 for the alignment procedures.

An example of a typical LEP2 physics event, Figure 2.5 shows the hadronic decay of a pair of W bosons, $e^+e^- \rightarrow W^+W^-$, where it can be seen that nearly all of the reconstructed tracks have two matching microvertex detector hits. All but two of the tracks are consistent with a single primary vertex. The remaining two tracks originate from a clearly visible neutral particle decay, showing the capability of reconstructing secondary vertices.

Vertex Detector

The vertex detector (CV) [46, 47] is a 1 m long and 0.47 m diameter cylindrical drift chamber located within the pressure bell, between the outer beampipe and the jet chamber (cf. Figure 2.6). As mentioned in the previous section, it is used to determine the position of decay vertices of short-lived particles and to improve momentum resolution. The chamber is segmented radially, and has an inner layer of 36 axial sectors and an outer layer of 36 stereo sectors.

In the inner layer, each sector has a plane of 12 anode or sensing wires strung parallel to the beam direction, located between radii of 10.3 and 16.2 cm with a spacing of 5.3 mm. The outer layer of stereo sectors each has 6 anode wires lying between radii of 18.8 and 21.3 cm with a spacing of 5 mm. Each anode wire defines a drift cell. Ionization from a charged particle passing through the cell is collected at the anode wire. The anode wires are 20 μm diameter gold plated W-Rh, the potential wires are 200 μm diameter gold plated Cu-Be and the cathode planes use 125 μm diameter Cu-Be wires spaced 1 mm apart. The cathode planes form the boundaries between adjacent sectors.

The stereo angle of the wires in the outer layer is $\sim 4°$. Information from these wires is used to obtain precise information about the z coordinate of the particle trajectory. All of the anode wires are offset from the plane defined by the potential wires by an alternating stagger of $\pm$ 41 μm in order to resolve the ambiguity as to whether a particle passed to the right or left side of the wire plane. The average spatial resolution of the chamber is about 55 μm. See Figure 2.7 for a schematic of the CV wire layout.

The axial cells provide a precise measurement of position in the $x-y$ plane by accurately measuring the arrival time of the electrons at the anode wire. The maximum drift distance of about 1.5 cm limits the effects due to diffusion. A good multi-hit detection capability is necessary

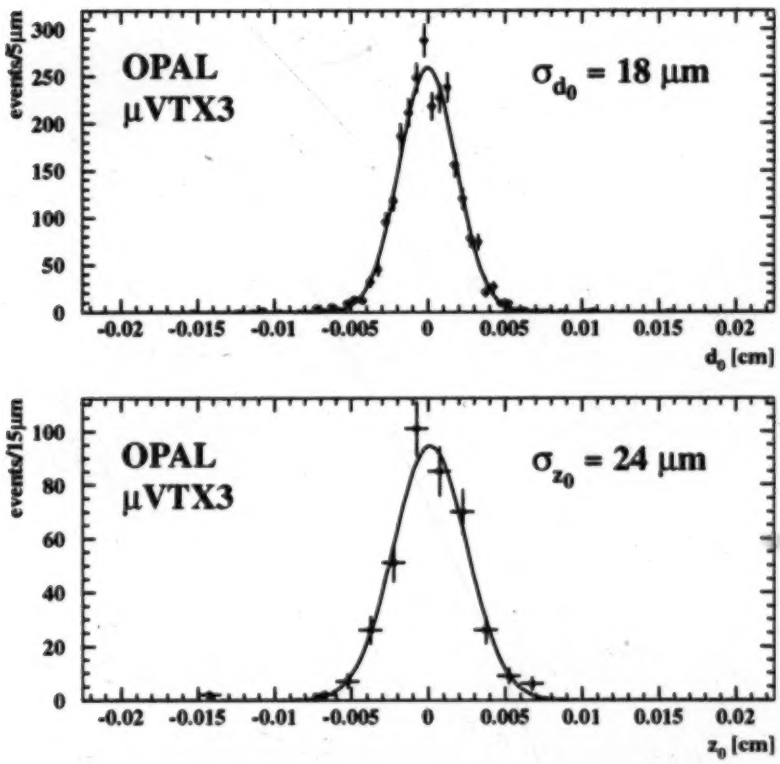


Figure 2.4: Impact parameter distributions in the $r-\phi$ plane (d_0) and along the beam axis (z_0) for $e^+e^- \to \ell^+\ell^-$ ($\ell = e, \mu$) events taken in 1996. The points with error bars are the data and the full line is a Gaussian fit [45].

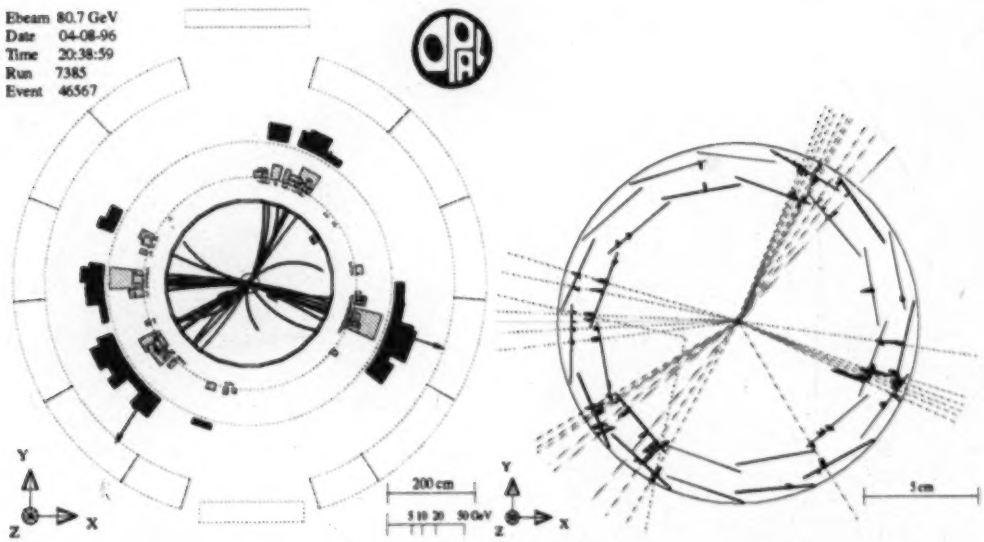


Figure 2.5: Example of an $e^+e^- \to W^+W^- \to q\bar{q}\,q\bar{q}$ event at LEP2. The left image presents the charged tracks and energy deposits reconstructed for the whole detector. The right image shows the hits in the SI microvertex detector, indicated by thick bars, which start at the position of the hit and have a length proportional to the signal height.

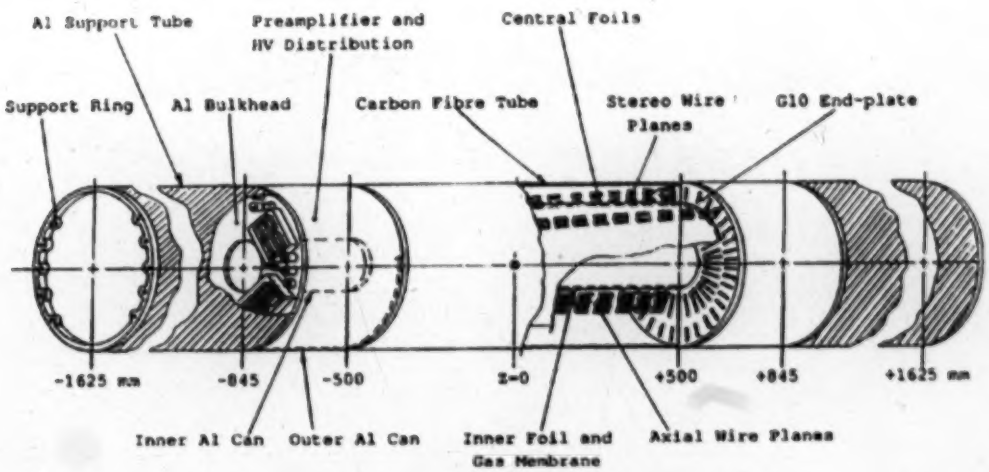


Figure 2.6: Cut-away view of the Vertex detector.

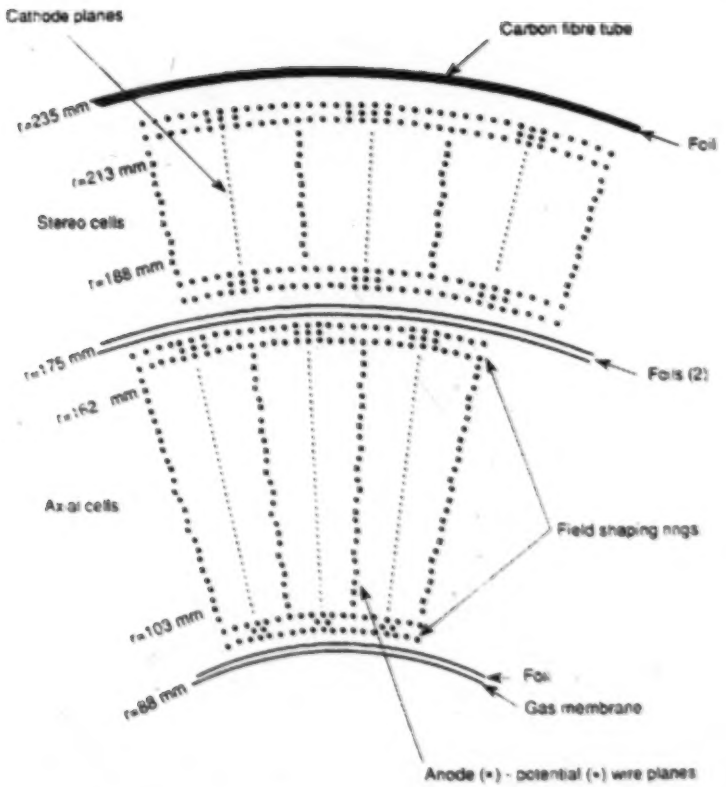


Figure 2.7: Endplate wire layout for a section of the Vertex detector.

to resolve individual particles within jets. By reducing the discrimination time between successive pulses to about 40 - 50 ns, a two particle separation of about 2 mm is achieved, assuming an average electron drift velocity of about 40 μm/ns in a drift cell.

The signals from each wire end are then passed to a unit which measures the difference in their arrival times. A measurement of this difference to 0.1 ns allows the z coordinate to be determined to roughly 4 cm. This information is used in the trigger and in offline track finding and fitting. A more precise z measurement is obtained by combining axial and stereo cell information in the offline tracking analysis.

Jet Chamber

As it is of critical importance to the study of hadronic jets, in this section we will not only discuss the general parameters of the jet chamber (CJ) but present more a more detailed performance review of the detector.

Like the vertex detector, the jet chamber [48, 49, 50] is of cylindrical design and surrounds CV within the pressure bell. CJ records the tracks of charged particles and is designed to combine good space and double track resolution with accurate momenta determination through measurement of their curvature in the magnetic field along with the possibility of particle identification by measuring dE/dx through multiple sampling of the energy loss in the gas. A track traveling perpendicularly to the anode wires encounters 4.9% X_0 of material due to chamber gas.

The chamber is 4 metres long with an inner radius of 25 cm and an outer radius of 185 cm. It is divided into 24 identical sectors each with 159 anode wires parallel to the beam line. As in the vertex chamber, cathode wire planes form the boundaries between adjacent sectors. The anode wires are grounded and lie between radii of 25.5 and 183.5 cm, equally spaced by 1 cm, alternating with potential wires. The maximum drift distance varies from 3 cm for the inner wires to 25 cm for the outermost wires. To resolve left-right ambiguities, the anode wires are staggered by ± 100 μm alternately to the left and right side of the plane defined by the potential wires.

Within the range $43^\circ < \theta < 137^\circ$, up to 159 points can be measured along each track, and at least eight points on a track are obtained over a solid angle of 98% of 4π. For each drift cell, the $r\phi$ coordinates are determined from the wire position and drift time. The integrated charge is measured at each end of the wire and the ratio determines the z coordinate to within several centimetres. The sum of the charges is used to calculate the energy loss, dE/dx, of the particle. The operating point of the jet chamber (gas gain $\sim 10^4$ and a drift field of 890 V/cm) is chosen as a trade-off between requirements for good dE/dx information (low gas gain) and good z resolution (high gas gain).

The jet chamber has had quite a successful run to date. The number of bad channels which are unavailable to track reconstruction, due to failures in the electronic readout chain, are typically kept as low as 0.3%. Single hit efficiencies of > 99% are measured by examination of the distribution of the number of hits on tracks in $e^+e^- \rightarrow Z^0 \rightarrow \mu^+\mu^-$ events. Of primary importance to the ability to reconstruct hadronic jets of closely packed tracks, Figure 2.8 shows the ability to resolve closely spaced hits, where 80% of hits greater than 2.5 mm are resolved. The difference in these efficiencies, for all hits and those where both hits are associated with tracks, is due to a low level of imperfect hit assignment by the pattern recognition algorithm [51].

Figure 2.9 presents the distribution of $1/p$ for μ^+ and μ^- tracks in $Z^0 \rightarrow \mu^+\mu^-$ events collected at the Z^0 peak during the 1998 data run. After suitable cuts, the momentum of each track should be equal to the beam energy, $E_{Beam} \approx 45.6$ GeV. A momentum resolution of $\sigma_p/p^2 = 1.5\times10^{-3}\ (GeV)^{-1}$ is measured compared to the design goal of $1.6 - 2.0\times10^{-3}\ (GeV)^{-1}$. Including a term due to multiple scattering, the momentum in the $r-\phi$ plane (p_t in GeV) is measured with a resolution of

$$\frac{\sigma(p_t)}{p_t} = \left(0.02^2 + (0.0015 p_t)^2\right)^{\frac{1}{2}} .$$

The impact parameter resolution, σ_{d_0}, of CJ is determined by studying the "miss distance"

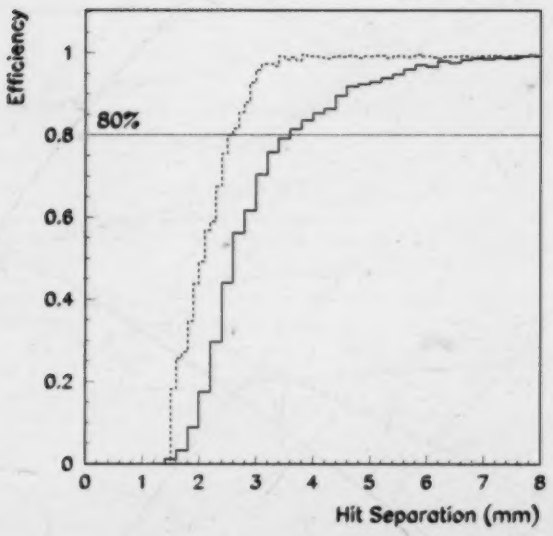


Figure 2.8: This plot, from the 1992 run, shows the efficiency of resolving a double hit (dashed line) and a double hit associated to a track (solid line) as a function of the double hit separation in the jet chamber. A double hit is resolved with 80% efficiency at a double hit separation of 2.5 mm. A double hit associated to a track is resolved with 80% efficiency at a double hit separation of 3.5 mm.

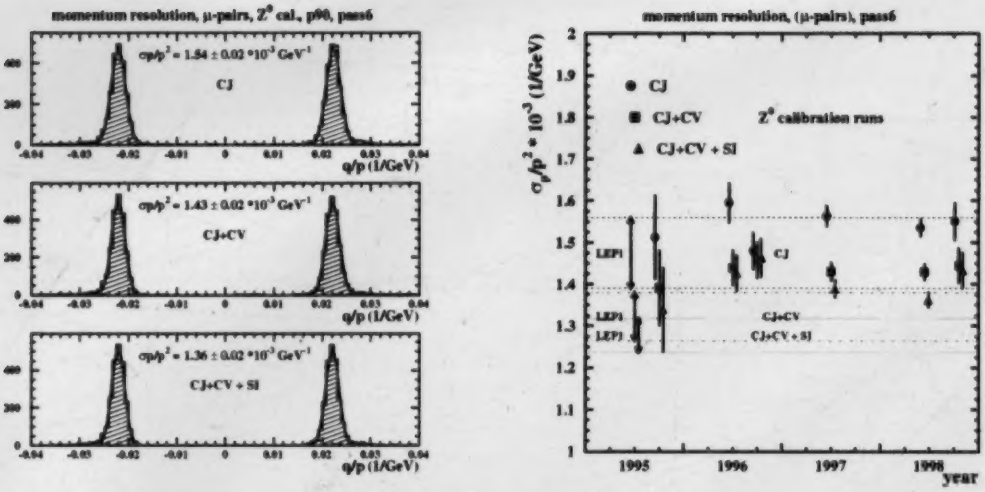


Figure 2.9: The plot on the left shows the distribution of $1/p$ for μ^+ and μ^- tracks in $Z^0 \to \mu^+\mu^-$ events collected at the Z^0 peak during the 1998 data run. The first subplot presents the measurement using only the jet chamber (CJ), then with improvements through inclusion of additional track information provided by the vertex chamber (CV) and the silicon microvertex detector (SI) respectively. The plot on the right shows the summary of the momentum resolution throughout the 1995-1998 data taking periods.

between muon tracks extrapolated to the beam axis in the $r-\phi$ plane using $Z^0 \to \mu^+\mu^-$ events. Figure 2.10 shows the intrinsic $r-\phi$ resolution determined using isolated $e^+e^- \to (Z/\gamma)^* \to \mu^+\mu^-$ events collected during the 1998 run. For the jet chamber alone, this distribution is centred on zero with a resolution of 98 μm. In general, the impact parameter of a track is not determined by the jet chamber alone. Including the additional track information from CV and SI, the overall impact parameter resolution is measured to be 19 μm.

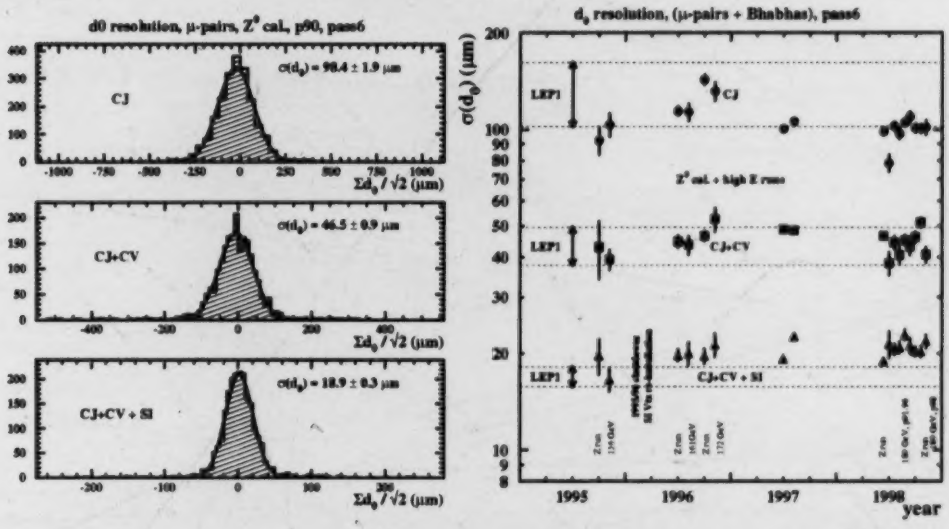


Figure 2.10: The plot on the left shows the instrinsic coordinate resolution, σ_{d_0}, measured using μ^+ and μ^- tracks in $Z^0 \to \mu^+\mu^-$ events collected at the Z^0 peak during the 1998 data run. The first subplot presents the measurement using only the jet chamber (CJ), then with improvements through inclusion of additional track information provided by the vertex chamber (CV) and the silicon microvertex detector (SI) respectively. The plot on the right shows the summary of σ_{d_0} throughout the 1995-1998 data taking periods.

The jet chamber does not perform as well in reconstructing the z coordinates of tracks. Using the method of charge division, which has been shown to be particularly sensitive to systematic effects due to asymmetric long range crosstalk induced on the signal wires, and the time difference of arrival times at each end of the signal wires a resolution of $\sigma_{z_0} = 1.6$ cm is achieved in hadronic Z^0 decay events and 2.5 cm for $Z^0 \to \mu^+\mu^-$ events. Including information from the vertex and z chambers, discussed below, this resolution improves to $\sigma_{z_0} = 1.0$ cm, which is dominated by the width of the beam.

The dE/dx resolution for $\mu^+\mu^-$ events [51] was calculated to be 3.1%. These values are somewhat better than the values which would be obtained from multihadronic events. Figure 2.11 shows dE/dx values, measured using the jet chamber, for multihadronic tracks and muon-pairs together with the expected functional form. In the region of the relativistic rise, a particle separation with a probability of at least 2σ is possible between electrons and pions for energies up to 13 GeV, and between pions and kaons or protons for energies up to 20 GeV.

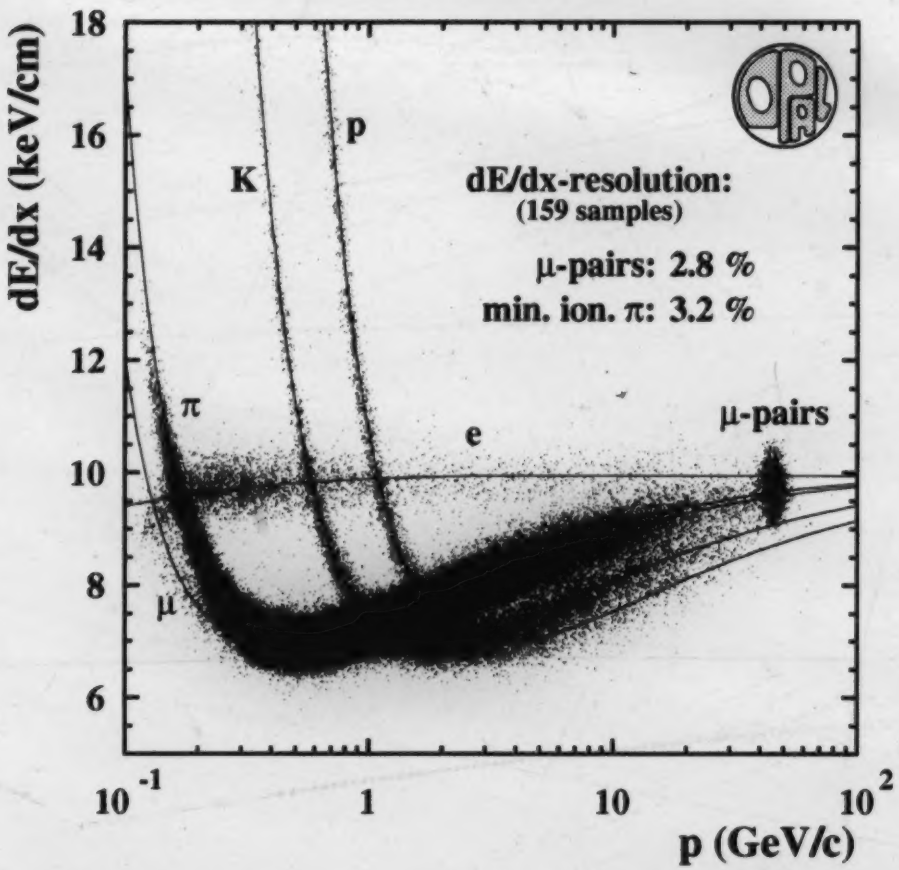


Figure 2.11: This large scatter plot shows the measured $\mathrm{d}E/\mathrm{d}x$ (1994 data) for multihadronic tracks and muon-pairs collected at the Z^0 peak, together with the expected functional form. The $\mathrm{d}E/\mathrm{d}x$ resolution for minimum ionising pions within multihadrons ($p = 0.4 - 0.8$ GeV) and muon-pairs with 159 $\mathrm{d}E/\mathrm{d}x$ hits is also indicated.

Z-Chambers

The Z-chambers (CZ) [52] are used to obtain a precise measurement of the z coordinate of charged particles as they leave the jet chamber, thus improving polar angle resolutions. The precise z measurements from these chambers improve both the polar angle and invariant mass resolutions, which are both extremely important in standard particle jet reconstruction schemes [53, 54].

The Z-chamber system consists of 24 panels, 4 metres long, 50 cm wide and 5.9 cm thick. Each panel is divided into eight 50 x 50 cm cells, giving a maximum drift distance of 25 cm in the z direction. They are arranged to form a barrel layer around the jet chamber covering the polar angle range $44° \leq \theta \leq 136°$ ($|\cos\theta| \leq 0.72$) and 94% of the azimuthal acceptance. The maximum drift voltage is about 20 kV giving a drift field of 800 V/cm. Each cell has six anode wires with 4 mm spacing and a stagger of $\pm 250\,\mu$m. Charge division is used to obtain the ϕ coordinate, in a method analogous to that used to determine the z coordinate in the jet chamber. The $r-\phi$ resolution is about 1.5 cm and the z resolution about $300\,\mu$m.

2.2.2 Time-of-Flight

The Time-of-Flight (TOF) systems consists of two detectors, the Time-of-Flight Barrel (TB) and Endcap (TE). The TOF system: generates trigger signals; helps in particle identification; and aides in rejection of background particles such as cosmic rays.

Time-of-Flight Barrel

The TB system covers $|\cos\theta| \leq 0.82$ and consists of 160 scintillation counters, each 6.4 m long forming a barrel of mean radius 2.36 m.

Time-of-Flight Endcap

In contrast to earlier LEP studies at the Z^0 resonance, the research programme at LEP2 is characterized by low cross-sections and an emphasis on new particle searches. Amongst the important signatures of new physics are final states with unbalanced visible momenta; events containing isolated photons; or those events with heavy, long lived (out of time) particles. In all cases, the performance of OPAL would benefit by paying particular attention to improving particle tagging efficiency, hermiticity and time resolution. This has been attained by the development of a new system of scintillating tiles which was installed in the endcap region of OPAL during the 1995-96 winter shutdown. Two layers of scintillating tiles, complementing the existing TB scintillators in the barrel region, were installed between the endcap presampler and the electromagnetic calorimeter (see Figure 2.12): the Tile Endcap (TE), providing full azimuthal coverage over $0.85 \leq |\cos\theta| \leq 0.95$; and the Minimum Ionizing Particle (MIP) Plug, again providing full azimuthal coverage over 42 mrad $\leq \theta \leq$ 200 mrad. The University of Alberta played a central role in the design, construction and commissioning of the TE system [55].

Given the goal of improving tagging efficiency, hermiticity and time resolution mentioned above, the essential requirements for the scintillating tile arrays were that they provide a high and uniform light yield, low noise and a time resolution of ≤ 5 ns. A number of physical constraints were also imposed: there existed a limited ($\approx$ 20 mm) thickness available for the detector; a restricted volume of cabling which could exit the endcaps; the 0.4 T magnetic field; and the required solid angle coverage. The only practical solution was to use scintillating tiles with embedded wavelength shifting (WLS) fibres, which were chosen to optimize the collection of scintillation photons ($\lambda \sim$ 420 nm). There are a total of 120 TE and 32 MIP plug tiles in each endcap. All are 10 mm thick and vary in area from 470 to 900 cm^2. The scintillation light collected by the WLS fibres is re-emitted at a longer wavelength ($\lambda \sim 500$ nm) and routed, by clear fibres, outside the detector where the magnetic field is low enough for the use of conventional, shielded, photomutiplier tubes.

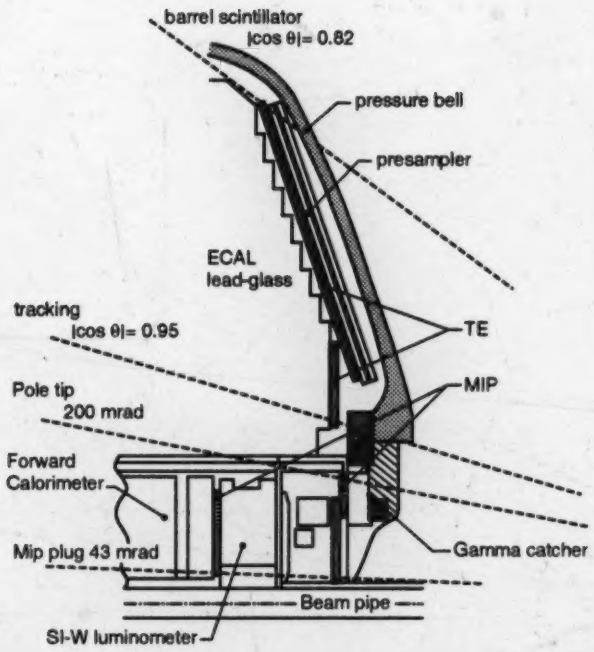


Figure 2.12: Presented above is the $y - z$ plane projection of the the position of the Tile Endcap and MIP plug in the OPAL detector. The dashed lines represent particle trajectories at the limits of the polar angle coverage by the various subdetectors. The dashed lines converge at the e^+e^- interaction point, approximately 2.5 m to the right of the centre of the figure.

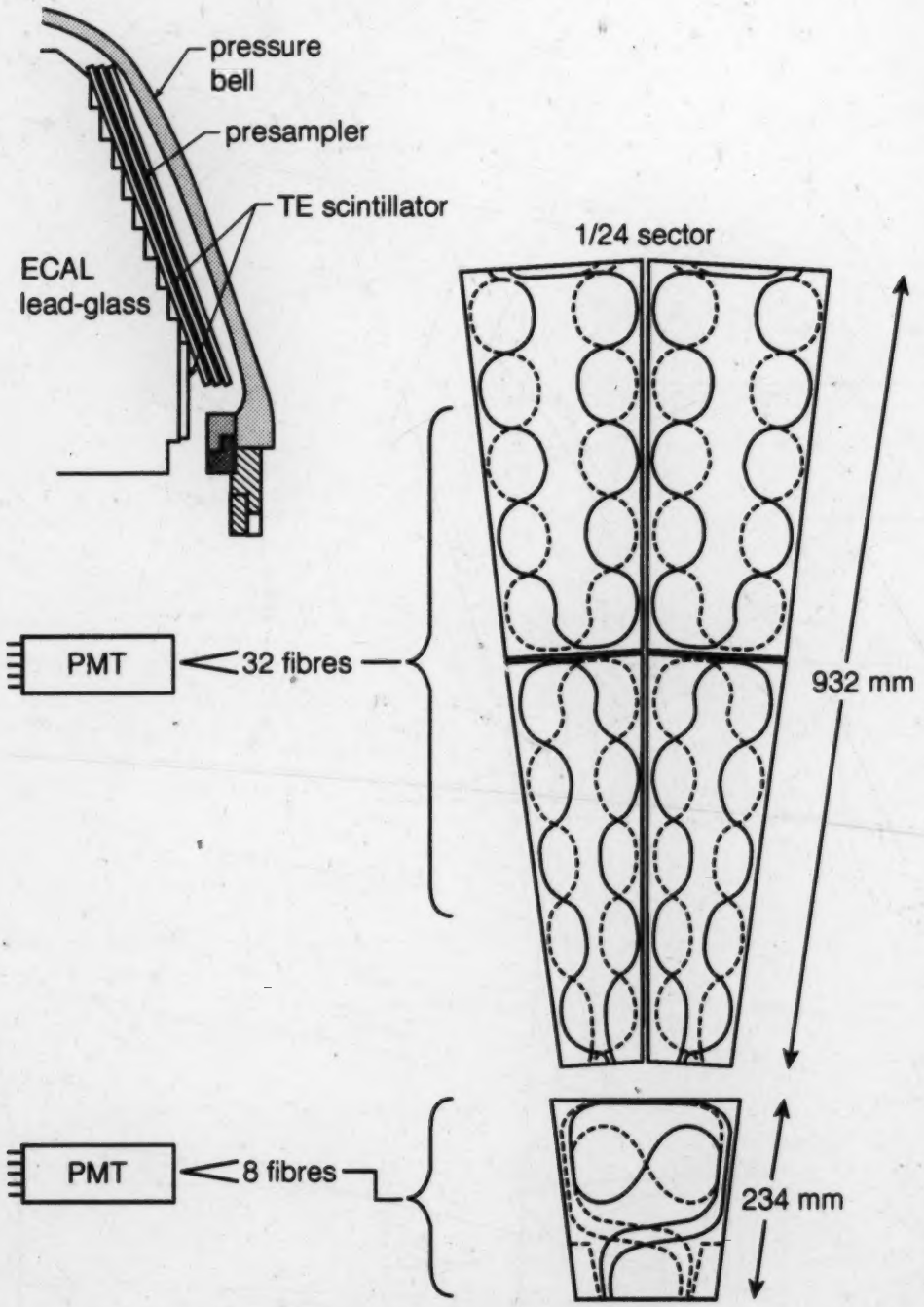


Figure 2.13: The Tile Endcap sector arrangement.

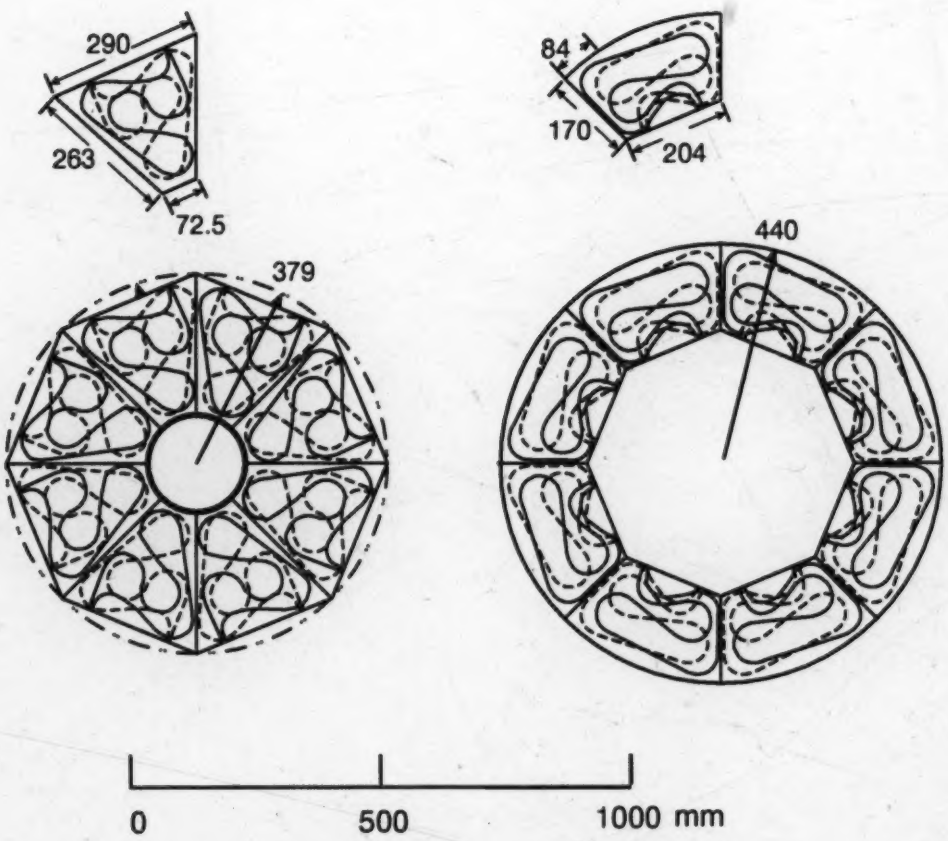


Figure 2.14: The MIP plug sector arrangement showing inner and outer layers.

The scintillating tiles in each endcap are arranged in sectors as shown in Figures 2.13 and 2.14. The outer two annuli of TE were subdivided into 48 sectors in order to match the 24-fold azimuthal segmentation of the OPAL trigger logic while having at least 2 separate scintillator elements per sector to match the truncated cone shape of the endcap electromagnetic calorimeter. To economize on photomultiplier channels, the light output from tiles in the same outer 1/24 sector were grouped together on the same photomultiplier. The inner 1/24 sectors are single piece tiles. Each of these is read out by separate photomultipliers, allowing for disconnection of the inner sectors from the trigger in case of high background rates caused by their proximity to the beamline. For the MIP Plug, the tile layout was dictated by: the need to cover the required angular regions with at least a two-fold coincidence of approximately annular layers; the azimuthal subdivision needed for trigger and veto functions; and by the high beam-induced backgrounds expected. The tiles are arranged as an inner and an outer pair of annular layers, each divided into 8 tile sectors.

The precise timing information provided by TE simplifies the unambiguous determination of the collision time, t_0. Figure 2.15 shows the bunchlet identification as seen by the MIP plug and the TE tiles. The observed spacing between bunches is 335 ± 1.0 ns, compared to the expected value of 334 ns. Gaussian fits to the bunchlet peaks yield a time jitter of 3.0 ns in bunchlet arrival times compared to the design requirements of 5 ns. This precise determination of t_0 is used throughout the OPAL subdetectors to provide more robust and selective track triggers in order to veto out-of-time hits such as those caused by cosmic rays, as well as by new particle searches looking for long lived, heavy neutral particles. The TE scintillators, situated behind the presampler material, as shown in Figure 2.12, may be the only source of the t_0 if the products are forward-going neutrals.

2.2.3 Calorimetery

Electromagnetic Calorimeter

The electromagnetic calorimeter system measures the energies and positions of electrons, positrons and photons, ranging in energies from tens of MeV up to the beam energy. It is a total absorption calorimeter for electromagnetic showers and is mounted outside the pressure vessel of the central tracking system (and after the magnetic coil and the time of flight system), covering 98% of the solid angle. Two distinct systems make up the electromagnetic calorimeter: a system of preshower counters (presampler) and a lead-glass calorimeter. The material in front of the electromagnetic calorimeter is approximately 2 radiation lengths thick which ensures that most electromagnetic showers will have started before entering the calorimeter, necessitating the preshower detector.

The main component of the electromagnetic subdetector is the array of lead-glass blocks. Lead-glass was chosen for its excellent intrinsic energy resolution. The energy measurement of a photon or electron in lead-glass is based on collecting the Cerenkov light from the secondary electrons generated in a electromagnetic shower. The blocks are designed to have good optical transmission and to present a reasonably large number of radiation lengths to prevent shower leakage. The resolution depends primarily on the photo-counting statistics.

The lead-glass system coupled with the presampler system provides some π^0-photon discrimination and, in conjunction with the central tracking system, electron-hadron discrimination. The angular resolution of electromagnetic clusters is approximately 4 mrad both in θ and ϕ for shower energies above 10 GeV.

Presampler

The barrel presampler (PB) consists of a cylinder of limited streamer mode wire chambers between the TB system and the barrel lead-glass calorimeter. There are 16 chambers covering the surface of a cylinder of radius 2.4 m and a length of 6.62 m. Each chamber is 3 cm thick and consists of two layers of limited streamer mode tubes with sense wires running axially. Readout is obtained by detecting signals induced on the 1 cm wide cathode strips located on both sides of each layer of tubes and oriented at 45° to the wire direction. The strips on opposite sides of a layer are

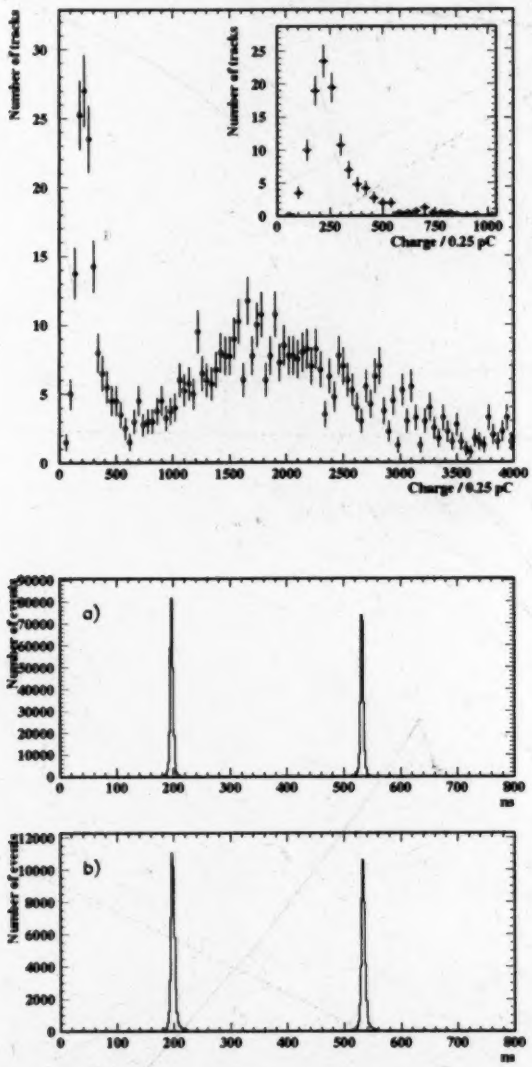


Figure 2.15: The uppermost plot shows the TE response to single tracks, corrected for angle of incidence, summed over all segments. The insert shows the same spectrum for the minimim ionizing particles, selected with the additional constraint that the associated electromagnetic cluster energy be less than 3 GeV. The lower plot shows the bunchlet identification as seen by the MIP plug tiles (a) and the TE tiles (b). The observed spacing bewteen bunches is 335 ± 1.0 ns, compared to the expected value of 334 ns. Gaussian fits to the bunchlet peaks yield a time jitter of 3.0 ns compared to the design requirements of 5 ns.

orthogonal. In addition, the charge collected on each wire is measured at both ends to provide a z position. Each module contains 480 pairs of strips in each orientation, and 192 wires.

The resolution for the position of electromagnetic showers, in the plane perpendicular to the shower direction, varies from about 6 to 4 cm as the shower energy varies over 6 to 50 GeV. This corresponds to an angular resolution for the photon trajectories of approximately 2 mrad. The resolution in z from current division is $\approx$ 10 cm for a single charged particle.

The endcap presampler (PE) is an umbrella shaped arrangement of 32 thin multiwire chambers, operating in high gain mode, in 16 sectors located between the pressure bell of the central tracking system and the endcap calorimeter. They cover the full azimuthal angle and the polar annular region defined by $0.83 < |\cos\theta| < 0.95$. The endcap presampler has an angular resolution of $\sigma_{PE} \approx 4.6$ mrad.

The anode wire plane consists of 432 (69) gold plated tungsten wires of 50 μm diameter spaced by 2 mm. The cathode read-out is done via 32 (22) strips and 16 (10) pads; the numbers in parenthesis refer to the small chamber. The total thickness of the chamber is 6.8 mm, and the width between the two cathode planes is only 3.2 mm.

Lead-Glass Calorimeter

The barrel lead glass calorimeter (EB) consists of a cylindrical array of 9440 lead-glass blocks of 24.6 X_0, located at a radius of 2455 mm, outside the magnetic coil covering the full azimuthal angle and the polar annular region defined by $|\cos\theta| < 0.82$. The calorimeter is segmented into 59 blocks in the z direction and 160 blocks in the ϕ direction. The longitudinal axes of the blocks are arranged such that each block points not at, but approximately near to (approximately 30 mm) the interaction point. This prevents neutral particles from escaping through gaps in between the blocks. The blocks are instrumented with magnetic field tolerant phototubes. Each lead-glass block is $\approx 10 \times 10$ cm^2 in cross-section and 37cm deep. The blocks are polished and, for optical isolation, each is wrapped in a sheet of vinyl flouride, the inner surface of which is coated with aluminum to improve the collection of Cerenkov light.

The endcap lead glass calorimeter (EE) consists of two dome-shaped arrays, each containing 1132 lead-glass blocks, located between the pressure bell of the central tracking system and the pole tip hadron calorimeter. EE covers the full azimuthal angle and the polar annular region defined by $0.81 < |\cos\theta| < 0.98$. Because of tight geometrical constraints, the lead-glass blocks are mounted with their axes coaxial with the beam line. The detector follows the curve of the pressure bell and thus the blocks come in three lengths, typically $\sim 22\ X_0$.

The spatial resolution of electromagnetic showers was found to be approximately 5 mm for EB and 2-5 mm for EE. The resolution for minimum ionizing particles is better and has been found to be approximately 2 mm and 2-4 mm, respectively. The energy resolution of EB has been measured to be $\sigma_E/E = (1.2 \pm 0.3)\% + (20.8 \pm 1.8)\%/\sqrt{E}$, and $\sigma_E/E = (4.6 \pm 0.7)\% + (12.5 \pm 3.6)\%/\sqrt{E}$ for EE. In Figure 2.16 we see the energy resolution of the EB calorimeter determined for photons in $\ell^+\ell^-\gamma$ events collected at the Z^0 peak [56].

Hadron Calorimeter

The hadronic calorimetry system measures the energy of hadrons which emerge from the electromagnetic calorimeter and assists in muon identification [2]. It is built in 3 sections: the barrel, the endcaps and the poletips. The iron used in the hadronic calorimeter provides at least four interaction lengths over a solid angle of 97% of 4π, as well as providing the flux return of the magnetic field. This yoke is segmented into layers, with planes of detectors between each layer. Since hadronic interactions may occur in the 2.2 interaction lengths of material preceding the hadron calorimeter, it is necessary to use information from the electromagnetic calorimeters to correctly measure the hadronic energy.

[2]Most hadrons are absorbed in the hadron calorimeter, leaving only muons to pass through.

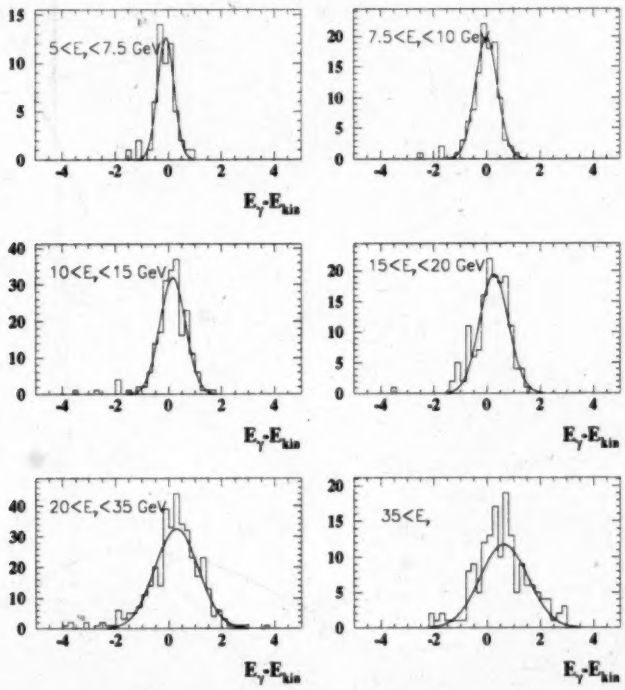


Figure 2.16: Electromagnetic calorimeter energy resolution, in GeV, of photons determined from $\ell^+\ell^-\gamma$ events collected at the Z^0 peak.

The barrel part of the hadronic calorimeter (HB) consists of nine active layers, alternating with eight, 10 cm thick, iron slabs together forming a cylinder between radii of 3.39 and 4.39 metres. The barrel ends are closed off by toroidal endcap regions (HE) consisting of eight layers of chambers alternating with seven layers of iron.

The active layers are limit-streamer mode cells with pad and strip readout. The streamer tube devices are strung with anode wires, 1 cm apart, in a mixture of 75% isobutane and 25% argon. The HB strips have an axial geometry and a width of 10 mm, giving a ϕ resolution of about 2.5 mrad, which is sufficient for muon identification.

The signals from the wires are used only for monitoring purposes and the energy measurement is obtained by analysing the induced charge collected on pads located on the outer chamber surface and from aluminum strips located on the inner surface. In the barrel, the strips are parallel to the beam line and in the endcaps they are perpendicular.

For fast trigger development, the layers of pads are grouped together to form towers which divide the solid angle into 48 bins of 7.5° in ϕ and 21 bins in θ. The signals from the pads are summed to provide an estimate of the energy in the shower. The 57,000 strips are used to provide muon tracking information and for shower profile mapping.

The pole-tip region (HP) extends the coverage of the hadron calorimeter from $|\cos\theta| = 0.91$ down to 0.99. In this region the number of sampling layers is increased to 10. The detectors are 0.7 cm thick multiwire proportional chambers containing a gas mixture of 55% CO_2 and 45% n-pentane. As in HB and HE, the chambers have pads on one side and strips on the other.

The hadron calorimeter is sufficiently thick so as to reduce the amount of *punch through* (particles from hadronic showers which penetrate the calorimeter volume) to less than 0.8% of all particles at 5 GeV. The response of the hadronic calorimeter, at fixed angle, was shown to be [35] linear with energy and the energy resolution was measured to be $\sigma_E/E = 120\%/\sqrt{E}$ for pions with an incident energy of 10 GeV.

2.2.4 Muon Detector

The muon detector is a system of large-area drift chambers constructed as a barrel, measuring r, ϕ and roughly z; and two endcaps, measuring x, y as well as z by the surveyed detector position. It covers the iron yoke almost completely; 93% of the solid angle is covered by at least one layer of the detector. The amount of material that a particle has to traverse before reaching the muon detector exceeds the equivalent of 1.4 m of iron (over 7 interaction lengths for pions). This reduces the probability of a pion not interacting in the hadronic calorimeter, thus faking a muon, to less than 0.001. The efficiency for detecting isolated muons above 3 GeV is essentially 100%.

The barrel part of the detector (MB) contains 110 drift chambers and covers $|\cos\theta| < 0.68$ with four layers and $|\cos\theta| < 0.72$ with one or more layers. Each cell is 0.6 m wide, 1.5 cm thick and the majority of length 10.4 m. The drift chamber hits have a resolution of 1.5 mm in the $r - \phi$ coordinate, corresponding to an angular resolution of 0.3 mrad. The z coordinate of each hit is given by a combination of several pad readouts yielding a resolution of 2 mm.

The endcap detector (ME) covers $0.67 < |\cos\theta| < 0.98$ with four layers of limited streamer tubes which are perpendicular to the beam axis. Using a series of 10 mm strip readouts, hits in each layer can be reconstructed, permitting the measurement of both x and y coordinates for particles traversing the chambers, with an angular resolution similar to that of MB.

Muon identification relies on extrapolating the track seen in the central tracking system through the iron absorber, allowing for energy loss and multiple coulomb scattering, then looking for the track in the muon detector which matches in position and angle in two views. The positional and angular accuracy required for the track measurement are determined by the multiple scattering of the highest energy muons of interest and are about 2 mm and 3 mrad respectively.

2.2.5 Forward Detectors

In new particle searches, 4π detector hermiticity is always a desirable goal, but just as critical is a precise measurement of the machine luminosity which is needed in order to set the most accurate exclusion limits possible. The forward detectors (FD) are used to measure the machine luminosity by detecting the small-angle Bhabha scattering events, as well as providing instrumented coverage as close to the beam pipe as possible, maximizing the detector acceptance. The forward detector system consists of; from 1993 onwards, the silicon-tungsten calorimeter; and most recently the MIP plug, combining to push the instrumented coverage of OPAL to $\theta > 43$ mrad.

Silicon Tungsten Calorimeter

The silicon-tungsten detector (SW) consists of two cylindrical small-angle calorimeters, encircling the beampipe at about ± 2.4 m from the interaction point, have an angular acceptance of $25 < \theta < 59$ mrad. Each calorimeter is built up from 19 layers of silicon sampling wafers and 18 layers of tungsten, corresponding to a total of 22 radiation lengths. The radial resolution on electron showers is approximately $10\mu m$ with an energy resolution of $\sigma_E/E \sim 28\%/\sqrt{E}$.

Forward Calorimeter

The forward calorimeter (FK) has 35 sampling layers of lead-scintillator sandwich, divided into 16 azimuthal segments, presenting 24 radiation lengths, readout with wavelength shifter to vacuum phototetrodes. It is divided into a presampler of 4 X_0 and the main calorimeter of 20 X_0. The

presampler has a wavelength shifter on the outside, but the main calorimeter is readout on both the inner and outer edges, allowing a θ measurement. The energy resolution has been measured as $\sigma_E/E \sim 18\%/\sqrt{E}$. The polar angle resolution on electron showers is ± 4 mrad near the inner edge, but degrades to ± 10 mrad at the outer edge. The azimuthal resolution, determined from the ratio of signal in adjacent segments, is approximately $\pm 2°$.

The Minimum Ionizing Particle Plug

The newest addition to the forward detectors, the Minimum Ionizing Particle (MIP) Plug, is described in Section 2.2.2.

Tube Chambers

Between the presampler and the main sections of the calorimeter there are three planes of brass-walled proportional tube chambers (FB). Individual shower positions have polar angle resolution of ± 2 mrad. Within the tube chamber acceptance the tube chambers and calorimeter provide combined cluster information. The energy measurement comes from the calorimeter which also provides θ and ϕ for showers falling outside the tube chamber acceptance.

Gamma Catcher

The gamma catcher (FE) is a small electromagnetic calorimeter physically extending in θ from 143 to 193 mrad. It thus covers the annular region between the outer physical edge of FK and the inner EE lead-glass calorimeter. It comprises eight independent azimuthal segments at each end, which provide a coarse determination of ϕ. Each segment consists of a lead-scintillator sandwich which is read out along the outer edge with a wavelength shifter bar and two 1 cm^2 silicon photodiodes. The calorimeter is non-containing with approximately 7 X_0 of active material, preceded by further 3 X_0. A high energy electromagnetic shower will thus be shared either with FK or the inner EE blocks depending on the θ angle of the interacting particle. The energy resolution of Bhabhas is $\approx 20\%$.

Far Forward Monitor

The far forward luminosity monitor (FF) counters are small lead-scintillator calorimeters, approximately 50 mm $\times$ 150 mm $\times$ 20 X_0 is size, mounted on either side of the beam pipe at 7.85 m from the primary interaction point. They are used for high statistics, realtime, luminosity measurements and to monitor beam backgrounds.

2.2.6 The OPAL Trigger and Data Stream

OPAL only records events if they satisfy certain trigger conditions, allowing one to separate interesting physics processes from uninteresting background processes, detector noise and beam-crossings where no interaction occurs.

Triggers

Subdetector trigger signals are of two types, 'stand alone' signals (high thresholds) and signals from $\theta - \phi$ binning (low thresholds). The high threshold trigger signals are typically multiplicity counts or energy sums. The low threshold trigger signals are formed from one of the $6(\theta) \times 24(\phi)$ overlapping bins which cover the 4π solid angle of the detector. The programmable trigger logic makes a decision by forming spatial correlations between subdetectors in $\theta - \phi$ along with stand-alone signals.

The central trigger [57] logic is installed in a dedicated Eurocrate with a special 'trigger bus' in addition the standard VME bus. Logical combinations of the signals on the trigger bus (*i.e.*

subdetector stand-alone signals and $\theta - \phi$ matrix outputs) are formed by the patter arrangement module (PAM), which uses look-up tables to derive the trigger decision from 120 possible outputs.

The global trigger unit (GTU) sends its trigger decision to the local trigger units (LTUs) of each subdetector. If the trigger decision is negative, a reset pulse is distributed 6 μs before the next bunch crossing. If the decision is positive the GTU transfers a central event number and other information to the LTUs. Each LTU inhibits further triggers when subdetector readout is in progress. The amount of time during the readout when no new triggers can be taken is called deadtime.

Multihadronic Trigger

CV and CJ provide fast measurements of the $r - z$ coordinates of hits in bins of θ and ϕ. Tracks are identified as straight lines in $r - z$, thus enabling a track based trigger decision. The time-of-flight counters, outside the magnetic coil, provide trigger signals in 21 bins of ϕ. The EB and EE calorimeters together are divided into $6(\theta) \times 24(\phi)$ overlapping bins and triggers are formed from analogue energy sums in each bin.

Some trigger conditions relevant for multihadronic events (any of which alone will cause a trigger) are:

- > 3 tracks
- > 2 tracks in the barrel region, $|\cos\theta| \leq 0.82$
- track trigger-TOF coincidence in ϕ
- track trigger-ECAL coincidence in $\theta - \phi$
- TOF-ECAL coincidence in ϕ
- ≥ 7 TOF hits in ϕ
- ≥ 7 GeV total energy in EB
- ≥ 6 GeV total energy in EE
- ≥ 2.6 GeV energy in a $\theta - \phi$ bin EB
- ≥ 3 GeV energy in a $\theta - \phi$ bin EE

Data Stream

Each subdetector is read out separately by its own front-end electronics to the local system create (LSC). This digitized information is collected and merged into a single data structure by the event builder (EVB) VME system. Each complete event is passed to the filter where events are checked, analysed, monitored and compressed before being written to disk. Further event selection can occur at this stage by rejecting obvious background events. Information generated from each event is copied from the filter disk to the ROPE processor farm, where the full OPAL reconstruction code ROPE [3] is run and where the calibration constants are applied.

[3] Reconstruction of OPAL Physics Events

Chapter 3

Artificial Neural Networks

When signal efficiency is a concern, as it is when a process of interest has a cross-section many times smaller than those of the associated backgrounds, optimally, we would prefer a method where all the parameters relevant to an analysis can be taken into account to determine the likelihood that a particular set of values came from a signal-like event. The Artificial Neural Network (ANN) is one such method.

Our ability to learn from observation is our primary source of knowledge about the world. We learn to classify objects, to tell cats from dogs or an 'A' from a 'B', on the basis of instances presented to us, not by being given a set of classification rules. Experience also teaches us how to predict events, such as a coming rain storm, and to estimate unseen quantities, such as when we judge the likely weight of an object from its size and appearance. Without this ability to learn from empirical data, we would be unable to function in daily life.

Theories and methodologies of learning are interesting from a number of perspectives. Psychologists try to model learning abilities of animals and to formulate high-level theories of how learning operates, while neurobiologists try to understand the biological mechanisms of learning at a lower level. Workers in artificial intelligence would like to understand in a more general way how learning is possible in a computational system, and engineers try to apply such insights to produce useful devices. Philosophers would like to understand the fundamental nature of inductive learning. Statisticians develop methods of inference from data that for certain tasks are more sensitive and reliable than unaided common sense. It is in the latter sense that we will consider neural networks. The applications of interest in this thesis are those designed for predicting an unknown category on the basis of known attributes.

Although neural networks were originally intended as abstract models of the brain, this biological connection is well developed in many references [58] and will not be explored in this thesis. Instead, we will attempt to develop an understanding of the neural network as a strictly mathematical idea.

To set the scene, it is useful to begin with a definition of what is meant by 'Neural Net'. A Neural Network [58] is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The processing ability of the network is stored in the inter-unit connection strengths, or weights, obtained by a process of adaptation to, or learning from, a set of training patterns.

An artificial neural network (ANN) consists of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections. The architecture of the ANNs considered here are based on very similar building blocks which perform the processing. These are variations on the parallel distributed processing (PDP) idea and lend themselves well to numerous applications in High Energy Physics (HEP), especially the growing interest in network based computing due to its inherit speed and scalability.

The logical components of an ANN can be listed as

- A set of processing units, known as *neurons*.

- Connections between the units. Generally, each connection is defined by a weight, w_{ij}, which determines the effect of the signal from unit j on another unit i.
- An external input, known as the *bias*, θ_i for each unit.
- A propagation rule, which determines the effective input, i_i, to unit i, from its external inputs.
- An activation function, $g_i(x)$, which determines the nonlinear response of the i^{th} unit based the effective input i_i to that unit.
- A state of activation, $a_i = g_i(x)$, for the unit, which is equivalent to the output of the unit.
- A learning rule. That is, the rule which determines the new level of activation of the unit based on its effective input and current level of activation.

In the following sections, a mathematical description of the basic neuron will be presented, followed by the construction of a neural network using a system of discrete neurons. The algorithm used to adapt the network state to solve a given classification task will be developed and issues relating to the implementation of neural networks in real world problems will be discussed. To begin, we will take a small tour through the history of neural nets [58].

3.1 A Brief History of Neural Nets

Ideas concerning machines that incorporate neural features have existed since work on the first general purpose computers began. In fact, the analogy between computing and the operation of the brain was to the fore in much of the early work in this area.

In 1942 Norbert Weiner [59] and his colleagues were formulating the ideas that were later christened 'Cybernetics', which he defined as 'control and communication in the animal and the machine'. Central to this programme, as the description suggests, is the idea that biological mechanisms can be treated from an engineering and mathematical perspective. Of central importance here is the idea of feedback.

In the same year that Weiner was formulating Cybernetics, McCulloch and Pitts [60] published the first formal treatment of artificial neural networks. The main result in this historic paper is that any well defined input-output relation can be implemented in a formal neural network.

One of the key attributes of networks is that they can learn from their experience in a training environment. In 1949, Donald Hebb [61] indicated a mechanism whereby this may occur in real animal brains. Essentially, synaptic strengths change so as to reinforce any simultaneous correspondence of activity levels between the pre-synaptic and post-synaptic neurons. Translated into the language of artificial neural nets, the weight on an input should be augmented to reflect the correlation between the input and the unit's output. Learning schemes based on this 'Hebb rule' have always played a prominent role.

The next milestone was the invention of the *perceptron* by Rosenblatt in 1957 [62]. One of the most significant results presented there was the proof that for a perceptron, a network consisting of a single input layer and a single output layer, a simple training procedure, known as the perceptron training rule, would converge if a solution to the problem existed.

In 1969 enthusiasm for neural networks was dampened somewhat when it was shown [63] that there exists an interesting class of problems (those which are not linearly separable) which single layer perceptron nets cannot solve. Further, it was felt there was little hope for the training of multi-layer systems which might deal successfully with some of these classification tasks. The fundamental obstacle to be overcome is the so-called 'credit assignment problem': in a multilayer system, how much does each unit (especially one not in the output layer) contribute to the error the net has made in processing the current training vector?

In spite of this theoretical setback, much work continued in what was now an unfashionable area, and there have been several developments in recent years that have led to a resurgence of

interest in nets. A breakthrough occurred in connection with so called *feed-forward* networks, when it was shown that the credit assignment problem had an exact solution. The resulting algorithm, 'Back error propagation' or simply Backpropagation, was popularized by Rumelhart, Hinton and Williams [64].

In order to investigate these new models, it is often necessary to resort to computer simulation, and the power of this method has obviously been greatly enhanced by advances in technology. Certain experiments would simply have been unthinkable fifteen years ago by the average worker, given the accessibility and power of the computing resources available. In recent years, there has been a marked increase in neural net research, both because of new mathematical insights and the enormous improvement in the ability to perform simulation.

It has been a little more than ten years since the first investigations into the application of neural networks in HEP were undertaken [65, 66] and much of todays work is still exploratory; however, the growth in applications to HEP is striking. The proof-of-concept for using ANNs in HEP [67] can be established by existing applications in off-line analysis of data from existing experiments. In the domain of pattern recognition; quark-gluon separation [68], heavy quark tagging [69] and new particle searches [70], results have been achieved with feed-forward networks which are superior to conventional approaches. For tasks such as mass reconstruction of a decayed particle from calorimetry information or other cases of functional approximation, feed-forward networks have outperformed standard methods [71]. Also, solving the optimization problem of track finding through the use of ANNs have been proposed [66] and successfully applied with real data [72]. The fast track reconstruction provided by ANN methods will prove to be crucial to the high speed event triggers [73] in the high luminosity environment of the future hadron collider experiments.

3.2 The Fundamentals

3.2.1 The Neuron

As mentioned previously, the basic logical unit of a neural network is known as a neuron. A neuron is a processing unit which has a number of inputs and one output.

Following observations of biological systems, a neuron performs a relatively simple job: it receives inputs from external sources and combines them, along with internal parameters, to form a highly non-linear response signal which can then be propagated to other neurons downstream.

This functionality is captured in the artificial neuron known as the Threshold Logic Unit (TLU), originally proposed by McCulloch and Pitts [60]. While conceptually simple, the TLU model captures the main features seen in its biological counterpart. The i^{th} TLU neuron receives N inputs, a_j, from its $j = 1 \ldots N$ external sources through connections weighted by w_{ij}. These weights can have both positive (excitory) and negative (inhibitory) values. This information is combined in a weighted sum with a *bias*, or *threshold*, θ_i, to form the total input to the i^{th} neuron,

$$i_i = \sum_j w_{ij} a_j + \theta_i \; . \tag{3.1}$$

This linear input is then fed through a non-linear *transfer*, or *activation*, function $g(i_i)$, to produce the non-linear response of the neuron, a_i. That is,

$$a_i = g(i_i) = g(\sum_j w_{ij} a_j + \theta_i) \; . \tag{3.2}$$

This process is depicted schematically in Figure 3.1.

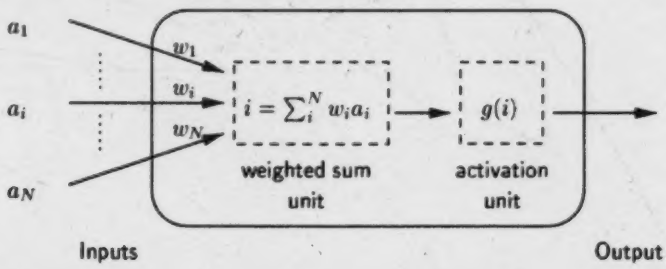


Figure 3.1: The structure of the basic neuron is shown. The unit collects a number of inputs, a_i, and generates their effective input, i, as a sum over the connections weighted by w_i. The activation, a, of the unit is then calculated by applying the transfer function, $g(x)$, to the effective input.

3.2.2 The Activation Function

McCulloch and Pitts originally proposed,

$$g(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{3.3}$$

known as the *Threshold* or *Heaviside* function which modelled the "on/off" property observed in the biological systems. However, further research in the field lead to the preference for differentiable, non-decreasing transfer functions. The differentiable, linear extension of the Heaviside function is the *Threshold Linear function*,

$$g(x) = \begin{cases} 1 & x \geq \frac{1}{c} \\ 0 & x < 0 \\ cx & \text{otherwise} \end{cases} \tag{3.4}$$

where c is a positive scaling constant. The activation derivative is:

$$g'(x) = \frac{dg}{dx} = \begin{cases} 0 & \text{if} \quad x \in (-\infty, 0) \cup [1/c, \infty) \\ c & \text{otherwise} \end{cases} \tag{3.5}$$

While differentiable, this function is not commonly used. More elegant solutions exist in a class of smooth, readily evaluated, functions with simple derivatives collectively known as *sigmoidal* functions.

The most used sigmoidal transfer function is the *logistic signal* function,

$$g(x) = \frac{1}{1 + e^{-\frac{x}{T}}} \quad \in (0, 1) \; \forall \, x \in \Re \tag{3.6}$$

which is sigmoidal and strictly increases for positive *temperature* constant $T > 0$. T sets the inverse gain of the function: a low temperature corresponds to a very steep sigmoid and a high temperature corresponds to an approximately constant function. The limit $T \to 0$ yields the Heaviside function, Equation 3.3. Strict monotonicity implies that the activation derivative of g is positive:

$$g' = \frac{dg}{dx} = \frac{1}{T} g(1 - g) > 0 \, . \tag{3.7}$$

Another common function in use is the hyperbolic-tangent function,

$$g(x) = \tanh(\frac{x}{T}) = \frac{e^{\frac{x}{T}} - e^{-\frac{x}{T}}}{e^{\frac{x}{T}} + e^{-\frac{x}{T}}} \quad \in (-1, +1) \forall \, x \in \Re \tag{3.8}$$

where $T > 0$ is a positive scaling constant. Again, g is monotone increasing ($g < 1$), so the activation derivative of g is positive,

$$g' = \frac{dg}{dx} = \frac{1}{T}(1 - g^2) > 0 \tag{3.9}$$

The choice of activation function can change the behaviour of a neural network considerably. For hidden units, the standard choice is to use a sigmoidal type function, either in symmetric, $(-1, 1)$, or asymmetric, $(0, 1)$, form. Sigmoidal functions are global in the sense that they divide the parameter space represented by the inputs into two halves: one where the response is approaching 1 and another region where it approaches 0 (-1). Hence, it is very efficient at making sweeping cuts in feature space. Typical High Energy Physics applications are most concerned with pattern classification and so these type of divisions in feature space are extremely beneficial.

For output units, the standard choice for classification tasks is again to use sigmoidal functions. The outputs are typically trained on binary target values, 0 or 1, depending on which class the input training pattern was assigned. However, for continuous transfer functions the realized output values, will generally not be exactly equal to one of these target values. One is still able to interpret these outputs in a very useful way, if the following sufficient conditions are met:

- The training is accurate.
- The task is coded such that only one output unit is "on" at a time.
- A mean square error function is used.
- The training data patterns are selected with the correct *a priori* probabilities.

In this case the outputs correspond to Bayesian *a posteriori* probabilities [74], and we should find that a sum over the values of the output nodes would yield $\sum_{i=1}^{N} o_i = 1$. Due to the difficulty in proving all the sufficient conditions listed above exist, this sum is typically not identically one. To facilitate a winner-takes-all classification for networks with more than one output node, we ensure the unit sum by weighting the i^{th} output by a sum over all outputs, $o_i / \sum_j o_j$, in the case $o_i \in (0, 1)$ and $e^{o_i} / \sum_j e^{o_j}$, when $o_i \in (-1, 1)$, then taking the maximum weighted output as the classification.

In this analysis, the logistic signal, Equation 3.6, was used as the transfer function for all nodes. The use of Equation 3.8 was investigated but did not significantly affect the performance of the networks.

3.2.3 Building a Network

Exploring the obvious utility of the basic neuron, a vast number of network topologies can be imagined. When the output of one neuron is allowed to go to the input of any neuron, including itself, the options are almost limitless. However, the ultimate goal is to create a network topology which is amenable to the development of an algorithm which defines how the network can learn; be it pattern classification or functional representation. In this case, there are a number of well known network topologies which can be used.

One of the most common topologies used in High Energy Physics applications is the *feed-forward* network. Networks with this architecture are built by grouping a number of neurons in layers and connecting the outputs of neurons from one layer to the inputs of the neurons in the next, or downstream, layer (cf. Figure 3.2). There are no connections between nodes in the same layer, and output from nodes are only used by the immediate downstream layer. Within neural networks it is helpful to distinguish three types of neurons: *input* units which receive data from outside the system, and typically do not perform any processing; *Output* units which hold the network result for analysis outside the system; and *hidden* units, whose inputs and outputs remain internal to the system. That is, inputs are presented to the input layer, and the network response is generated by progressively feeding the output of one layer as inputs to the next, until the output layer

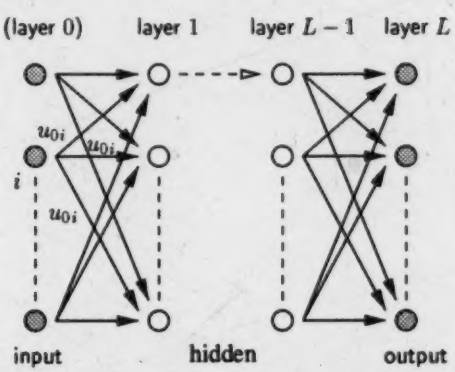


Figure 3.2: The general layout of a neural network. Layer 0 distributes the input to the input layer 1. The output of the network is (generally) the output of the output layer L.

is reached. This is called the *forward pass* phase of a feed-forward network analysis, for obvious reasons. However, without a method for adjusting the internal parameters of the network to achieve the desired response to some known set of input patterns, this method hardly seems interesting.

3.3 Learning with Neural Networks: The Delta Rule

Unlike conventional data processing techniques which require complex programming, neural networks develop their own solutions to problems. In effect, neural networks are trained rather than programmed.

A neural network has to be configured such that application of a set of inputs produces the desired set of outputs. Various methods to set the strength of the connections exist. This process is known as *training* the network and is accomplished by feeding the network training patterns and letting it adjust its internal state according to some rule.

We can categorize two distinct types of learning situations: *supervised* learning, in which the network is trained by providing it with both the input and the desired output patterns; and *unsupervised* learning, in which an output unit is trained to respond to clusters of pattern within the input. In this case, the system in supposed to determine the statistically salient features of the input patterns for itself. Unlike the supervised learning method, there are no *a priori* set of categories into which the patterns are to be classified. This is at odds with the envisioned application in this thesis, as we do have well defined categories in which we would like to classify the data. Further, this method is more sensitive to the relative population of signal events in the set of training patterns and as we have no *a priori* knowledge of the signal cross-sections, this would introduce an unavoidable bias. In this thesis, supervised learning was used to train all networks.

Virtually all learning rules for models of this type can be considered as a variant of the *Hebbian* learning rule [61]. The basic idea is that if two units, i and j, are simultaneously active, their interconnection should be strengthened. If unit i receives input from unit j, the simplest version of Hebbian learning prescribes that we modify the weight w_{ij} by,

$$\Delta w_{ij} = \gamma a_i a_j \ , \tag{3.10}$$

where γ is a positive constant of proportionality representing the *learning rate*. Another common rule uses not the actual activation of the i^{th} unit, a_i, but its error. That is, the difference between

the actual value and the desired, or *target*, value t_i,

$$\Delta w_{ij} = \gamma(a_i - t_i)a_j \ . \tag{3.11}$$

This is often called the *delta rule*.

It should be noted that the delta rule as stated in the previous equation only makes sense for an output node where the target value is known. How do we update the weight of a hidden unit? This is the credit assignment problem discussed in the introduction to this section, where it was mentioned that with feed-forward nets, it can be shown that the credit assignment problem can be solved by backpropagation [58, 64].

3.3.1 Backpropagation

Backpropagation is probably the most well researched training algorithm in neural network research and forms the starting point for most applications. The central idea to this solution of the credit assignment problem is that the errors for the units in the hidden layers are determined by back-propagating the errors of the units in the output layer. Backpropagation can be considered a generalization of the delta rule for non-linear activation functions and multi-layer networks. The discussion presented in this section is adapted from Ref. [75].

We begin by defining what we mean by network error. Formally for each training pattern, p, we assign a network error, E^p. This error function incorporates the actual and target values of the output nodes for the current pattern and is a measure of the degree to which the network has failed to generate the desired response. The goal, then, is to apply the specified learning algorithm to adjust the internal network parameters in such as way as to minimize the value of E^p over the whole training set.

The form of the error function is immaterial to the present discussion, we only require that it be a differentiable function of the values of the output units. The output units use differentiable activation functions which, in turn, are functions of the inputs and weights from the previous layer. Thus, through application of the chain rule, we see how we will develop the back-propagation algorithm. For clarity, however, the most common measure of the network error is the *Least Mean Square* measure. For a network with N_o units in the output layer, the error for pattern p is

$$E^p = \frac{1}{2}\sum_{i=1}^{N_o}(o_i^p - t_i^p)^2 \ , \tag{3.12}$$

where we have introduced the notation $o_i \equiv a_i$ for the activation of the i^{th} output unit. The network error is then defined as the sum of the error over all training patterns,

$$E = \sum_p E^p = \sum_p \left(\frac{1}{2}\sum_{i=1}^{N_o}(o_i^p - t_i^p)^2\right) \ . \tag{3.13}$$

The activation is a differentiable function of the total input, given by

$$a_i^p = g_i(i_i^p) \ , \tag{3.14}$$

in which

$$i_i^p = \sum_j w_{ij}a_j^p + \theta_i \ . \tag{3.15}$$

The transfer function $g(x)$ is indexed to indicate it need not be common to all units in the network, although this is typically the case. In order to minimize the error function, we make the change in the weights proportional to the negative of the derivative of the error, as measured on the current pattern, with respect to each weight. This method, known as *gradient descent*, thus defines

$$\Delta^p w_{ij} = -\gamma\frac{\partial E^p}{\partial w_{ij}} \ , \tag{3.16}$$

where $\gamma > 0$ is a constant of proportionality controlling the step size of the updating increment. Expanding the partial derivative in terms of the effective inputs to each of the output units, this becomes

$$\frac{\partial E^p}{\partial w_{ij}} = \frac{\partial E^p}{\partial i_i^p} \frac{\partial i_i^p}{\partial w_{ij}} . \tag{3.17}$$

We can use Equation 3.15 to rewrite the second factor as

$$\frac{\partial i_i^p}{\partial w_{ij}} = a_j^p . \tag{3.18}$$

If we now define

$$\delta_i^p = -\frac{\partial E^p}{\partial i_i^p}, \tag{3.19}$$

we will obtain the desired updating rule, resulting in a gradient descent of the error surface in weight space. Therefore, the weights are updated according to

$$\Delta^p w_{ij} = \gamma \delta_i^p a_j^p . \tag{3.20}$$

This is calculable for the output layer, where $\partial E^p / \partial i_i^p$ exists by definition of E^p, so the trick is to calculate the δ's for each hidden unit in the network. What the discoverers of backpropagation realized in the 1980's is that there exists a simple recursive algorithm to do just that, and this is implemented by propagating the error signals backward through the network.

Applying the chain rule, we write Equation 3.19 as

$$\delta_i^p = -\frac{\partial E^p}{\partial i_i^p} = -\frac{\partial E^p}{\partial a_i^p} \frac{\partial a_i^p}{\partial i_i^p} . \tag{3.21}$$

The first factor reflects the change in the error as a function of the output of the unit, and the second reflects the change in the output as a function of the change in its input. By Equation 3.14 we see that

$$\frac{\partial a_i^p}{\partial i_i^p} = g'_i(i_i^p) \tag{3.22}$$

which is simply the derivative of the transfer function of the i^{th} unit. We now understand the motivation for using transfer functions with easily expressed derivatives. Had this derivative merely existed, but not in an analytical form, significant computational time would be spent evaluating this derivative throughout the course of the training phase of the network.

To evaluate the first factor appearing in Equation 3.21 we will consider two cases. First assume that the i^{th} unit is an output unit of the network, written o_i. In this case, we have specified that $\partial E^p / \partial a_i^p$ should exist, thus applying the results of Equation 3.22 to Equation 3.21 yields

$$\delta_i^p = -\frac{\partial E^p}{\partial a_i^p} g'_i(i_i^p) \tag{3.23}$$

which is well defined for any output unit.

The second case to consider is when the unit is not an output unit. We will denote $h_i \equiv a_i$ as the activation of the i^{th} unit in the hidden layer with N_h units. In this case we do not readily know the contribution of the unit to the overall training error E but we do know E as a function of the inputs to the output layer, $E^p = E^p(i_1^p, \ldots, i_{N_o}^p)$, and we known the effect of each of the hidden units on each of the output units through Equation 3.15. We then use the chain rule to write

$$\frac{\partial E^p}{\partial h_i^p} = \sum_{j=1}^{N_o} \frac{\partial E^p}{\partial i_j^p} \frac{\partial i_j^p}{\partial h_i^p} = \sum_{h=1}^{N_o} \frac{\partial E^p}{\partial i_j^p} \frac{\partial}{\partial h_i^p} \sum_{k=1}^{N_h} w_{jk} h_k^p = \sum_{h=1}^{N_o} \frac{\partial E^p}{\partial i_j^p} w_{ji} = -\sum_{j=1}^{N_o} \delta_j^p w_{ji} . \tag{3.24}$$

Substituting this result into Equation 3.21 yields

$$\delta_i^p = g'_i(i_i^p) \sum_{j=1}^{N_o} \delta_j^p w_{ji}. \tag{3.25}$$

as the delta value for the i^{th} unit in the hidden layer, as a function of the delta values of those upstream units and the weights with which it is connected to them. With the delta's for the first hidden layer, this method can now clearly be applied recursively to generate the delta's for the whole network and, when applied in conjunction with Equation 3.16, defines the updating rule for the weights in the network. This procedure constitutes the generalized delta rule for a feed-forward network employing nonlinear units.

When the error function is given by Equation 3.13, as is used in the analyses presented in this thesis, the results derived above can be summarized in three equations:

- The weight of the connection is adjusted by an amount proportional to the product of an error measure, δ, on the unit i receiving the input, and the output of the unit j sending the signal along the connection,

$$\Delta_p w_{ij} = \gamma \delta_i^p a_i^p \,. \tag{3.26}$$

- If the unit is an output unit, the error measure is given by

$$\delta_i^p = -(o_i^p - t_i^p)\, g'_i(i_i^p) \,. \tag{3.27}$$

For the sigmoid activation function, Equation 3.6, $g' = \frac{1}{T} g(1-g)$ yields

$$\delta_i^p = -(o_i^p - t_i^p)\, \frac{o_i^p}{T}(1 - o_i^p) \,. \tag{3.28}$$

- The error measure for a hidden unit in layer L is determined recursively, in terms of error measures of the units to which it directly connects, and the weights of those connections. For the sigmoid transfer function,

$$\delta_i^p = g'_i(i_i^p) \sum_{j=1}^{N_{L+1}} \delta_h^p w_{ji} = \frac{1}{T} o_i^p (1 - o_i^p) \sum_{j=1}^{N_{L+1}} \delta_h^p w_{ji} \tag{3.29}$$

Improvements to Backpropagation: Learning Rate and Momentum

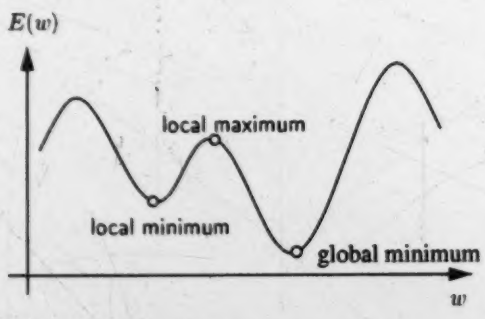


Figure 3.3: The total error as a surface in weight space.

The gradient descent learning procedure just developed requires that the change in the weight be proportional to $-\partial E^p/\partial w$. True gradient descent is based on infinitesimal steps through the weight space. This is computationally impractical so the algorithm implemented in Equation 3.16 employs a finite constant of proportionality, γ, to govern the steps sizes, or *learning rate*, in the incremental updating.

There are two concerns which determine the *a priori* size of γ. Suppose we start network training with some randomly set values of weights and thresholds. We perform a gradient descent to find the *global* minimum, Figure 3.3, of the global error function, Equation 3.13. However, there may exist a number of *local minima*, and during training the network may become trapped in the first minimum it encounters (which may or may not be the global minimum), corresponding to a partial solution of the network response to the training data set. If γ is too small, not only will the network take an inordinate amount of time to find a solution, but it may be unable to escape from even the most shallow local minima it encounters. Conversely, if γ is too large, the learning rate will be much faster but the network may oscillate wildly about the global minimum.

One solution to this problem is to redefine the updating algorithm such that each step in the weight space is governed not only by the error gradient but also by a random component which may flip the direction of the step (i.e., go up the gradient curve instead of down it), in the hopes that the network will be able to jostle itself out of a shallow local minima. Another method is to add a term to the updating algorithm which is proportional to the last updating step. This is known as a *momentum* term. The new rule replacing Equation 3.16 is

$$\Delta^p w_{ij}(n) = -\gamma \delta_i^p a_j^p + \alpha \Delta^p w_{ij}(n-1) \,, \tag{3.30}$$

where n indexes the current updating iteration and α is a constant which determines the contribution of the previous weight change. Thus if a previous weight change $\Delta w_{ij}(n-1)$ was large, so too will the next step $\Delta w_{ij}(n)$. That is, the weight change carries some momentum to the next iteration.

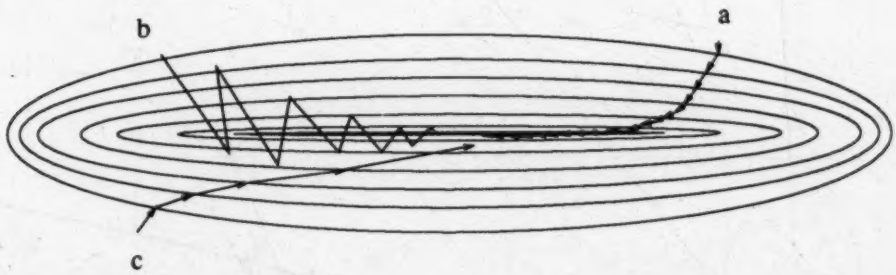


Figure 3.4: The gradient descent in weight space. Path a) is followed for a small learning rate, path b) for a large learning rate, showing large oscillations, and path c) for a large learning rate but with a momentum term added.

The role of the learning rate and momentum term can be seen in Figure 3.4. When no momentum term is used, it takes a number of steps for the minimum to be reached when using a low learning rate (small γ). With a faster learning rate (larger γ), the minimum is never reached because of the large oscillations. Adding the momentum term prevents the oscillations while allowing for a larger γ, and the minimum is reached must faster.

The crucial parameter in backpropagation is the learning rate, γ. The ideal choice for γ depends on two things: the scale of the activations and the *fan-in* of the network. If the average activation of the i^{th} node, $\langle a_i \rangle$, is large, the optimal γ for the weights connecting to that unit will be small.

The natural choice would be to rescale the inputs such that $\langle a_i \rangle \sim \mathcal{O}(1)$, in which case all units will have approximately the same optimal γ. The fan-in to a unit is the number of units connecting to that unit, in the forward direction. The optimal learning rate has been found to scale like $\gamma \propto 1/(\text{fan} - \text{in})$ [74]. The optimal value for the momentum term, α, is closely related to the choice of γ. α controls the averaging of the updates, and an increase in α translates to an increase in the effective learning rate. A choice of α in the range $0.5 < \alpha < 1$ is often helpful.

The feed-forward networks with backpropagation updating used in this thesis are defined by:

- The network topology: the number of units in the input layer, the number of hidden layers and the number of units in each of these, and the number of units in the output layer.
- The sigmoidal activation function used for each unit and the value of the corresponding temperature, T.
- The learning rate, γ, used in the updating steps, along with the contribution by the momentum term controlled by the value of α.

3.3.2 Performance Issues

With the development of a detailed approach to training a neural network in the last section, the actual implementation still needs to be discussed.

Generalization: A Measure of Network Performance

Depending on the structure of the neural network, for each problem and network architecture there exists a minimal *representation error*; this error is the error which is obtained with optimal weight and threshold parameters for that particular network and problem.

Finding optimal weight values means minimization of an error function which expresses the error in approximation by the neural network on a set of learning samples. This will lead to a low *optimization error*. However, it is important that the approximation represents the underlying data well, rather than being efficient at identifying random data patterns inherit to the specific sample training data set. A low *generalization error* is what is desired.

The generalization error is measured by testing the approximation over a separate data set, known as the *validation* set, obtained from the same process as the training sets. When the optimization and generalization errors lie far apart, the neural net is said to have *over-fitting*; the neural net has learned to represent the learning samples very well but does not represent the underlying process from which the training data was generated.

The *performance* of a neural network is its accuracy when presented with data other than that with which it has been trained. It must not be confused with its speed of operation, which is the conventional meaning of the term. The concept of performance is connected with generalization, such that a neural network that generalizes well will give good performance. Unfortunately, this is not always easy to attain. Good performance on the training set can often be achieved by building a large neural network and applying a large number of training cycles (many epochs). However, this does not guarantee good performance on unseen data. Indeed, a large complex neural network often gives very poor performance on previously unseen data such as that contained in the validation set. Achieving balanced performance involves optimizing the complexity of the neural network.

Neural Network Complexity

Neural network complexity is a subtle issue, which is not just a matter of the number of units used. It is, in fact, an indication of the modelling capabilities of a neural network: a complex neural network can better model the relationships inherent in a large training set, while a simple neural network cannot model the relationships as effectively. From Figure 3.5, we can infer the effect of

network complexity on the generalization error. A network with low complexity is unable to fit the trends in the data, resulting in both a large training error and generalization error. A network with medium complexity is better able to fit the trends in the data, resulting in reductions in both the training error and generalization error. Lastly, a network with high complexity begins to fit trends in the training data which is not a general feature of the process from which they were drawn. This results in a reduction in the training error but the generalization error increases.

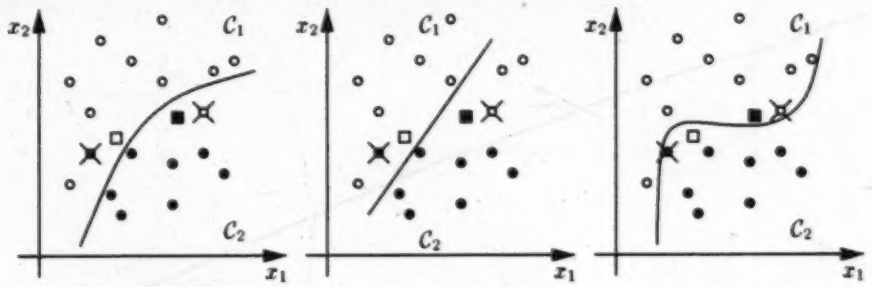


Figure 3.5: Network complexity as related to generalization error. The training set patterns are marked with circles, validation patterns are marked with squares, exceptions are marked with a ×.

Three main factors affect the complexity of a neural network and, hence, its ability to generalize. They are:

- The neural network size.
- The size of the training set.
- The number of training cycles, or epochs, used.

Changing the size of the network is the most obvious way of changing the complexity of a neural network. Unfortunately, the larger the neural network, the more likely it is to 'learn' any noise embedded in the training data, giving it a false significance which will lead to poor generalization. Good generalization is achieved by keeping the size of the neural network to a minimum compatible with the required level of performance. The optimum size is usually found by carrying out a number of training runs on differently sized networks and comparing their performance on validation data.

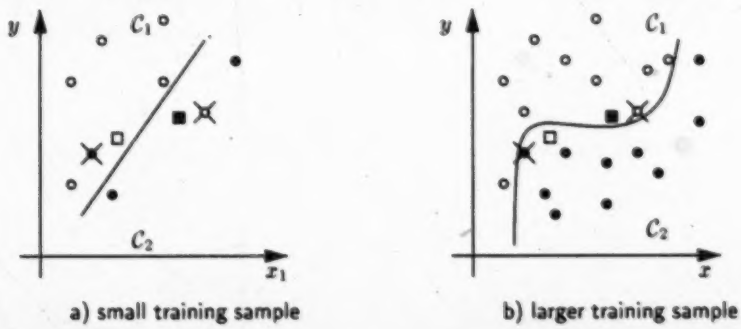


Figure 3.6: Network complexity as related to training sample size. The training set patterns are marked with circles, validation patterns are marked with squares, exceptions are marked with a ×.

Related to network size is the training sample size. This is easily understood by noting that a larger training sample size contains more information to fit the free parameters in the neural network model. Plot *a*) in Figure 3.6 shows what may happen when the training sample size is too small. The neural network was not presented with a sufficiently representative sample to efficiently categorize the two classes. In plot *b*) of the same figure we see the case were the training sample size was sufficiently representative of the parent classes to allow the network to identify the discriminating features.

The generalization performance depends mainly on the ratio N_w/N_p, where N_w is the numbers of weights in the network and N_p is the number of training patterns. For a feed-forward network with one hidden layer of sigmoidal thresholding neurons, the generalization error is of the order $\mathcal{O}(N_w/N_p)$ [76]. As a general rule, one should have at least ten times more training patterns than weights in the network. In High Energy Physics applications the amount of Monte Carlo data available for training is often limited, especially for the signal processes, so the only free parameter left to control network complexity is the network size. The process of removing neurons from the network is called *pruning* and will be discussed in detail in Section 3.4.

Given that the training sample size and network size are understood, at what stage do we call the network trained?

Early Stopping

When a neural network is first trained, it learns the gross statistical properties of the training set. As training proceeds through epoch after epoch, the network learns more of the detailed properties of the data. Unfortunately, training data is sometimes not perfect and may include noise that was present when the measurements were taken, and with further training, the network starts to model the noise in the training data. This results in over-training, leading to poor generalization and degraded performance with unseen data. One way to prevent this is to stop training at the point where the performance on unseen data starts to degrade. This is known as *early stopping*.

Early stopping is the preferred method used in this thesis. A network is trained over a large number of epochs and the resulting generalization error is observed. This error should fall initially, as the network begin to learns how to classify the input patterns. Over-fitting is observed when the generalization error ceases to fall and instead begins to rise again. This indicates the network has begun to classify the input patterns in the training set based on random internal correlations which are not inherit properties of the data set they were derived from. The final network state is chosen to be the point at which the generalization error on the validation set begins to rise. Figure 3.7 shows an example of the evolution of the training and generalization error. Although the training error is seen to continually fall, the network state is chosen at that point where the generalization error of the validation set begins to rise again.

Despite the demonstrated success of the backpropagation learning algorithm in many situations, there are some aspects which make the algorithm a less than optimal choice for all applications. Most troublesome is the long training process which can seemingly require disheartening amounts of training data and an inordinate amount of computing time. Some factors which affect the rate of network learning are discussed below.

Learning per Pattern

After the full set of learning patterns has been presented to the network for evaluation, it is said to have performed one *epoch*. Although, theoretically, the backpropagation algorithm performs gradient descent on the total error only if the weights are adjusted after each epoch, more often than not the learning rule is applied to smaller subsets of the training set, one at a time. There exists empirical indication that this results in faster convergence. Care must be taken, however, with the order that the patterns are taught. For example, when using the same sequence over and

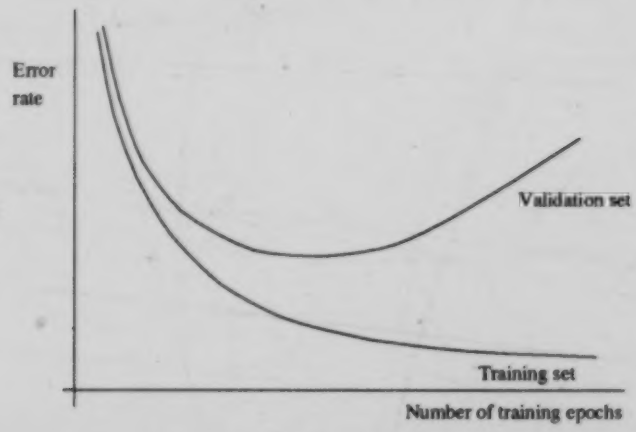


Figure 3.7: A representation of the evolution of the optimization and generalization error. Although the optimization error, as would be measured using the training set, is seen to continously fall, the generalization error, measured using the validation set, reaches a minimum then begins to rise indicating the onset of over-fitting.

over again, the network may become focused on the first few patterns. This problem is overcome randomly permuting the training set after each epoch.

In this thesis, the training sets were randomly permuted after each training cycle and the training networks were updated a number of times per epoch. The size of the subsets were selected so that, if the training set consists of N_C classes to be categorized, a random sampling could be expected to contain a representative of each class. This sample size was taken nominally to be at least $2N_C$, with larger values considered.

Network Saturation

Often long training times are indications of non-optimum network parameters: transfer function temperature, the learning rate γ or the momentum α. Much research has been devoted to developing algorithmic approaches to solving these issues but to date there are no clear solutions, and the tried-and-true method of brute force (change a parameter and observe the effect) is often still the best choice for ensuring an informed use of a neural network in an application.

Outright training failures generally arise in two ways: network paralysis and local minima. The trouble presented by the presence of local minima has already been discussed in Section 3.3.1, so we will concern ourselves with the issue of network paralysis. As a network trains, the weights can be adjusted to very large values which can lead to very large inputs being presented to the downstream neuron. If sigmoidal activation functions are used, this will result in an output very close to zero or one. Considering Equation 3.28, the weight adjustments will be very close to zero and the training process can be effectively halted. There are complex algorithmic methods which attempt to address this but are limited in scope and applicability. Without an algorithmic solution it is prudent to monitor for the occurrence of network paralysis.

The state of network training is captured by the networks *saturation* level. The saturation, s,

defined for each layer in the network, is given by,

$$s = \begin{cases} \frac{1}{N}\sum_i^N (1 - 2o_i^2) & \text{if } g(x) \in (0,1) \\ \frac{1}{N}\sum_i^N o_i^2 & \text{if } g(x) \in (-1,1) \end{cases}, \tag{3.31}$$

and measures the resolution of the units. An s value close to one signifies that the units in the layer are generating values very close to their extremes and implies that they have "made up their mind". An s close to zero means that the network is still learning and so s is a measure of to what extent the network is learning or not.

Pre and Post Processing

Input representation is crucial for achieving good performance. *Pre-processing* describes any process that converts inputs (signals or data) into a form suitable for use with a neural network. One form of pre-processing would be the application of a data reduction algorithm, such as a Principal Component Analysis, to decrease the number of input units, thus decreasing the number of weights in the network. Often times such dimensional reduction algorithms lose discriminating information which could be used by the network and so were not considered in this thesis.

Another form of pre-processing is to apply a non-linear scaling to the inputs so that they are all treated equally by the network. If one of the inputs is much larger in magnitude than the others, the network will concentrate on that one and take a very long time to learn the others. The network learns most easily if all inputs are scaled to $\mathcal{O}(1)$, where it can concentrate on the important parameter correlations rather than the saturation effects of a single input. Thus, the i inputs, x_i are shifted by their means, $\bar{x}_i$, and then scaled by their standard deviations σ_i, *ie.* $x_i \mapsto \frac{x_i - \bar{x}_i}{\sigma_i}$. The means and standard deviations were calculated using the Standard Model backgrounds only, correctly mixed by their relative cross-sections, which has the benefit that the results could be directly compared to data. Further, to avoid the possibility that the tails in some of these distributions could still be large, we transform as $x_i \mapsto \tanh\left(\frac{x_i - \bar{x}_i}{\sigma_i}\right) \in [-1, 1]$.

Post-processing, as it is related to the interpretation of the network output, was discussed previously.

3.4 An Algorithmic Approach to Pruning

Having already introduced many of the techniques used to ensure efficient application of the neural network method, we will return to the discussion of minimizing network complexity.

As stated previously, changing the size of the network is the most obvious way of changing the complexity of a neural network. Good generalization is achieved by keeping the size of the neural network to the minimum which is compatible with the required level of performance. The optimum size is usually found by carrying out a number of training runs on differently sized networks and comparing their performance on validation data. The process of removing neurons from the network is called *pruning*. Network pruning can be considered for two cases: pruning of the analysis variables and hence the number of units in the input layer; and pruning the number of units in the hidden layers.

3.4.1 Pruning of the Input Layer

The original pool of possible inputs to a neural network is selected by highlighting certain aspects of the signal or background processes, such as topology, event energy, *etc.* , which are felt to contain significant discriminating information.

Each input, however, incurs a cost in terms of added complexity to the network, which in turn requires larger data sets for efficient training [74]. Optimization of the ANN involves pruning of

these input nodes, as well as the nodes in the hidden layer, to remove those which do not contribute enough information to warrant their contribution to the complexity of the network.

A standard approach to selecting the relevant variables in an analysis is based on ordering their *figure of merit*, F [77], values. F measures the difference in distribution of a variable F for two classes of event; if $F \approx 0$, the distributions are similar and this variable has no separation power, while if $F = 1$ the distributions have no overlap and complete separation is possible. F is defined by the 'inverse correlation integral':

$$F = \alpha_1 \alpha_2 \int \frac{(f_1(x) - f_2(x))^2}{\alpha_1 f_1(x) + \alpha_2 f_2(x)} \, \mathrm{d}x \tag{3.32}$$

where class 1 contributes a fraction α_1 of the sample with normalized distribution $f_1(x)$, and class 2 contributes $\alpha_2 = 1 - \alpha_1$ with normalized distribution $f_2(x)$. This is performed intuitively in a standard cut based analysis.

If a variable has $F \approx 1$ between the signal and all backgrounds, it will be a useful parameter, in *both* standard cutting analysis as well as in a neural network based analysis. However, there may exist variables with $F \approx 0$ but which are highly correlated in such a way that this information is not easily observed in their projections. We would like a method which can realize these multidimensional correlations in a more automated manner. A multi-layer neural network attempts to separate regions of the input space into selected target classes (*ie.* signal and background), by constructing hyperplanes around these regions in the input hypervolume. Although not obvious in terms of their projections, and therefore in their F values, a neural network would have little difficulty separating two orthogonal normal distributions such as those presented in Figure 3.8. Other common pruning methods, such as the many variants of weight decay, also have their difficulties.

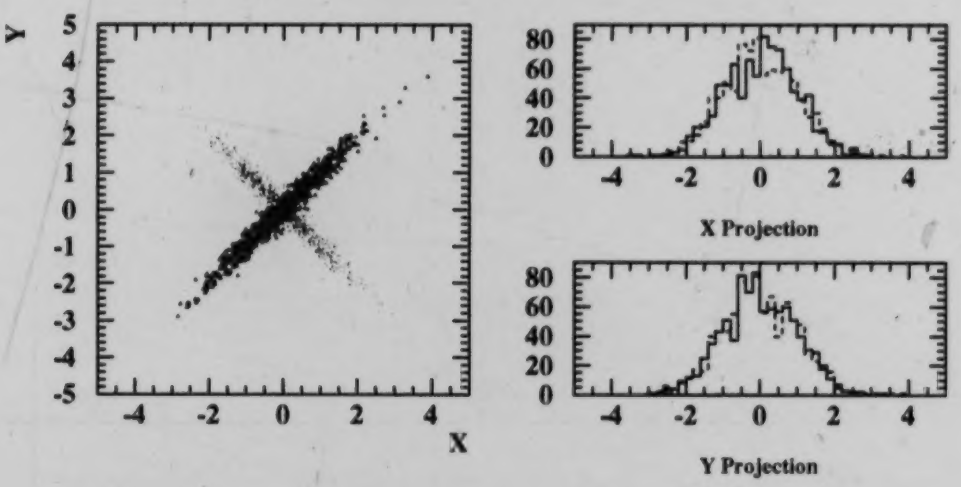


Figure 3.8: Shown are two variables, X, Y, where we assume their correlations are unknown to the researcher. These variables contain useful discriminatory information for a neural network analysis, but would be missed by a standard cut based analysis.

Instead, the most direct method to achieve the pruning of the input variables was used. The ANNs, distinguished by different sets of inputs, training parameters and topologies, were trained and their resulting ϵ_{signal} vs. expected background were graphically compared. In this way, one

could compare the results as well as monitor the stability of the network over various background sets and signal types. This is, of course, computationally intensive but in the applications common to HEP, the networks are relatively small and this approach can be accomplished quickly.

3.4.2 Pruning of the Hidden Layers

Pruning of the hidden layer presents a better opportunity for a more automated analysis, as one is not concerned with maintaining the aesthetically pleasing physical interpretations of the variables as with the input layer. If there are N *uncorrelated* inputs, the minimum number of hyperplanes needed to cut out a volume in the closed N-dimensional input space is $(N+1)$ [74]. Past this, the more hyperplanes one uses, the smoother the border defining the classes. If subsets of the inputs are *correlated*, this relation no longer holds and a smaller number of hyperplanes may be appropriate. The realization of these hyperplanes in the ANN topology is related to the number of hidden nodes.

When the inputs are correlated their contributions to the nodes in the hidden layer are often highly correlated and in the cases where these correlations persist in the *outputs* of the hidden layer nodes a type of Principal Component Analysis (PCA), common in Astrophysical papers [78, 79], can rotate the hidden layer output space to minimize these correlations. From this rotated space, a smaller number of variables can be found to contain the majority of the discriminatory information.

Below, we first remind the reader of the standard PCA technique, and then describe the variant of this method which was successfully used to prune the networks considered.

Principal Component Analysis

As an example, let us assume there exists a sample of P interesting events, where each event is defined by a N parameters which can be represented by a vector $\mathbf{x}^T = (x_1, x_2, \ldots, x_N)$. The goal of the standard PCA analysis is to concentrate the information stored in the N-dimensional points into a set of M-dimensional, $M < N$, points which are able to reproduce the original data set to a good approximation, in a least squares sense.

These events form a cloud of points in the N-dimensional parameter space, and we can imagine the case where there exists a direction, $\mathbf{a}_1$, along which this cloud is most elongated. The variation of the data is maximal in the direction of $\mathbf{a}_1$, and so contains the most useful information in terms of discriminating the data points. Collapsing the the data along $\mathbf{a}_1$ will yield a new data set of dimension $(N-1)$. This is the first order least squares description of the data and defines the first principal component. The i^{th} principle component is determined by successive iterations of this procedure, repeatedly collapsing the data along the 1^{st} to $(i-1)^{th}$ orthogonal basis vectors. This process can be shown [78] to be equivalent to finding the eigenvectors corresponding to the largest eigenvalues of the matrix,

$$\mathbf{C} = \sum_{i=1}^{P} (\mathbf{x}_p)(\mathbf{x}_p)^{\mathrm{T}}, \tag{3.33}$$

where $\mathbf{x}_p$ is the p^{th} sample vector. As C is clearly real and symmetric, the eigenvectors are obtained through diagonalization.

The maximum number of principle components which can be found in this manner is N. It may turn out, however, that the last $(N-M)$ principle components may contain minimal variation, thus contribute little to describing the data set. By describing the data set in terms of their M dimensional projections along the first M principle components, we will be able to account for most of the discriminatory information contained in the sample, but with a smaller parameter space. This dimensional reduction is often called *whitening* and is used in many image compression algorithms.

Optimization of PCA for Classification

The problem that needs to be addressed in our case is not simply how to compress the available information into a smaller subspace, but how to discriminate between different classes of events. The variant of PCA employed here [78] does not deconstruct a single set of inputs as above, but rather the set of difference vectors between events of different classes (again in a least squares sense).

We will denote $\mathbf{x}_k^{(\mu)}$ as the k^{th} sample vector of class μ. The job is to find a linear coordinate transformation, defined by a rotation matrix A and constant translation $\mathbf{c}$, such that the new coordinates, expressed as

$$\mathbf{y}_k^{\mu} = A^T(\mathbf{x}_k^{\mu} - \mathbf{c}) \,, \tag{3.34}$$

contain the maximum amount of discriminatory information based on a least-mean-square representation of all the difference vectors between the sets.

Let $\mathbf{a}_i$ be the normalized orthogonal column vectors comprising the rotation matrix A, $A = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_M)$, where $M \leq N$. We therefore seek the $\mathbf{a}_i$ and $\mathbf{c}$ which maximize the quantity,

$$S^2 = \sum_{\substack{k,l,\mu,\mu' \\ \mu \neq \mu'}} \left(\mathbf{y}_k^{(\mu)} - \mathbf{y}_l^{(\mu')}\right)^2 \,. \tag{3.35}$$

In order to determine the components of $\mathbf{c}$, we first appeal to the maximization principle of Equation 3.35. Rewriting it as,

$$S^2 = \sum_{\substack{k,l,\mu,\mu' \\ \mu \neq \mu'}} \left(A^T(\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')})\right)^2 \tag{3.36}$$

which when we maximize with respect to the constant translation $\mathbf{c}$, yields $\frac{\partial}{\partial \mathbf{c}} S^2 \equiv 0$, providing no information about $\mathbf{c}$. This tells us that the translation component of Equation 3.34 is not important in maximizing the inter-class variance. Rather, we note that once Equation 3.34 is determined, it will be applied on an event by event basis to assist in class categorization. In this sense, the rotation matrix contains all the discriminatory information; the translation $\mathbf{c}$ can then be determined by minimizing the residual error, E:

$$E^2 = \sum_{k,\mu} \mathbf{y}_k^{(\mu)^T} \mathbf{y}_k^{(\mu)} \tag{3.37}$$

$$\sum_{k,\mu} \left(\mathbf{x}_k^{(\mu)} - \mathbf{c}\right)^T AA^T \left(\mathbf{x}_k^{(\mu)} - \mathbf{c}\right) \tag{3.38}$$

$$\sum_{k,\mu} \left(\mathbf{x}_k^{(\mu)} - \mathbf{c}\right)^2 \,. \tag{3.39}$$

Minimizing E with respect to $\mathbf{c}$ we obtain,

$$\frac{\partial}{\partial \mathbf{c}} E^2 = \frac{\partial}{\partial \mathbf{c}} \left\{ \sum_{k,\mu} \left(\mathbf{x}_k^{(\mu)} - \mathbf{c}\right)^2 \right\} = 0 \tag{3.40}$$

which is satisfied when $\sum_{k,\mu}(\mathbf{x}_k^{(\mu)} - \mathbf{c}) = 0$, or

$$\mathbf{c} = \frac{1}{(N_\mu + N_{\mu'} + \ldots)} \sum_{k,\mu} \mathbf{x}_k^{(\mu)} \equiv \bar{\mathbf{x}} \,. \tag{3.41}$$

This result should not surprise us, since the mean spectrum should contain no discriminatory information so we simply choose the new origin to be the centre of gravity of the data sample before applying any rotation. For simplicity we now assume $\bar{\mathbf{x}}$ has been subtracted from the data, $\mathbf{x}^{(\mu)} \mapsto \mathbf{x}^{(\mu)} - \bar{\mathbf{x}}$.

With this transformation, we are now interested in finding the linear transformation,

$$\mathbf{y}_k^{\mu} = A^T \mathbf{x}_k^{\mu} , \tag{3.42}$$

which maximizes Equation 3.35.

Using Lagrange multipliers, we may write,

$$\frac{d}{d\mathbf{a}_i}\left[\sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(\mathbf{y}_k^{(\mu)} - \mathbf{y}_l^{(\mu')}\right)^{\mathrm{T}} \left(\mathbf{y}_k^{(\mu)} - \mathbf{y}_l^{(\mu')}\right) - \lambda_i \mathbf{a}_i^{\mathrm{T}} \mathbf{a}_i\right] = 0, \tag{3.43}$$

where λ_i is the Lagrange multiplier for $\mathbf{a}_i$. Substituting for $\mathbf{y}_k^{(\mu)}$ from Equation 3.42,

$$\frac{d}{d\mathbf{a}_i}\left[\sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')}\right)^{\mathrm{T}} \mathrm{A}\mathrm{A}^{\mathrm{T}} \left(\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')}\right) - \lambda_i \mathbf{a}_i^{\mathrm{T}} \mathbf{a}_i\right] = 0. \tag{3.44}$$

Defining $\mathbf{b}_{kl}^{(\mu\mu')} = (\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')})$, Equation 3.44 becomes,

$$\frac{d}{d\mathbf{a}_i}\left[\sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(\mathbf{b}_{kl}^{(\mu\mu')}\right)^{\mathrm{T}} \mathrm{A}\mathrm{A}^{\mathrm{T}}\, \mathbf{b}_{kl}^{(\mu\mu')} - \lambda_i \mathbf{a}_i^{\mathrm{T}} \mathbf{a}_i\right] = 0 , \tag{3.45}$$

or,

$$\sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \frac{d}{d\mathbf{a}_i}\left[\left(\mathbf{a}_1 \cdot \mathbf{b}_{kl}^{(\mu\mu')}, \ldots, \mathbf{a}_n \cdot \mathbf{b}_{kl}^{(\mu\mu')}\right) \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{b}_{kl}^{(\mu\mu')} \\ \vdots \\ \mathbf{a}_n \cdot \mathbf{b}_{kl}^{(\mu\mu')} \end{pmatrix} - \lambda_i \mathbf{a}_i^{\mathrm{T}} \mathbf{a}_i\right] = 0 . \tag{3.46}$$

which upon differentiation becomes,

$$2\left[\sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \mathbf{b}_{kl}^{(\mu\mu')} \left(\mathbf{b}_{kl}^{(\mu\mu')}\right)^{\mathrm{T}} - \lambda_i \mathrm{I}\right] \mathbf{a}_i = 0 , \tag{3.47}$$

or,

$$[\mathrm{C} - \lambda_i \mathrm{I}]\mathbf{a}_i = 0 . \tag{3.48}$$

Therefore the $\mathbf{a}_i$ are eigenvectors of the Hermitian matrix C:

$$\mathrm{C} = \sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')}\right)\left(\mathbf{x}_k^{(\mu)} - \mathbf{x}_l^{(\mu')}\right)^{\mathrm{T}} , \tag{3.49}$$

which is simply the covariance matrix of the difference vectors. The eigenvectors $\mathbf{a}_1, \ldots, \mathbf{a}_M$ of the covariance matrix define the linear transformation A in Equation 3.42.

Returning to Equation 3.35, we may express S^2 as,

$$S^2 = \sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(\mathbf{y}_k^{(\mu)} - \mathbf{y}_l^{(\mu')}\right)^2 = \sum_{\substack{k,l,\mu,\mu' \\ \mu\neq\mu'}} \left(A^T \mathbf{b}_{kl}^{(\mu\mu')}\right)^2 ,$$

$$
\begin{aligned}
&= \sum_{\substack{k,l,\mu,\mu' \\ \mu \neq \mu'}} \sum_i \left(A^T \mathbf{b}_{kl}^{(\mu\mu')} \right)_i^2 , \\
&= \sum_{\substack{k,l,\mu,\mu' \\ \mu \neq \mu'}} \sum_i \left(\sum_\rho A_{i\rho}^T (\mathbf{b}_{kl}^{(\mu\mu')})_\rho \right) \left(\sum_\sigma A_{i\sigma}^T (\mathbf{b}_{kl}^{(\mu\mu')})_\sigma \right) , \\
&= \sum_{\substack{k,l,\mu,\mu' \\ \mu \neq \mu'}} \sum_i \left(\sum_\rho \sum_\sigma A_{i\rho}^T (\mathbf{b}_{kl}^{(\mu\mu')})_\rho (\mathbf{b}_{kl}^{(\mu\mu')})_\sigma A_{\sigma i} \right) , \\
&= \sum_i \left(\sum_\rho \sum_\sigma A_{i\rho}^T C_{\rho\sigma} A_{\sigma i} \right) , \\
&= \mathrm{Tr}\left[A^T C A \right] = \sum_i \lambda_i , \qquad (3.50)
\end{aligned}
$$

where the last equality follows from the fact that the rotation matrix A diagonalizes C. Clearly, the larger the eigenvector, the larger the discriminating power of the corresponding eigenvector. Therefore, the M eigenvectors with the largest eigenvalues give the best M-dimensional linear discrimination between classes μ and μ'.

Principal Component Pruning: Application of PCA to Neural Networks

In supervised learning, a network is presented with a set of data vectors, $\mathbf{x}_k$, where k is the k^{th} input pattern, and an associated target output $\mathbf{t}_k$, $\dim(\mathbf{x}) = N$ and $\dim(\mathbf{t}) = M$. The success of the network classification is judged, in a least squares sense, by the ability of the network to minimize the generalized error,

$$E = \sum_k (\mathbf{o}_k - \mathbf{t}_k)^2 , \qquad (3.51)$$

where $\mathbf{o}_k$ is the observed network response to the input training patterns.

Large networks have more functional flexibility than smaller networks, so are better able to fit the training data. However large networks can have higher parameter variance than small networks, resulting in poor network generalization. The number of parameters in a network is a crucial factor in its ability to generalize.

No practical method exists for determining, *a priori*, the proper network size and connectivity. The most direct approach is simply to remove a hidden node, retrain the network and calculate the generalization error, repeating the process until the generalization error rises to an unacceptable level. This is obviously extremely time consuming.

A promising approach is to start with a large, fully connected network and remove the *information* passed through the network by projecting out the least significant principal components, as determined using the algorithm derived in the previous section. This method has the advantage that the *effect* of removing a hidden node can be determined without actually changing the network topology. Nodes can be removed one-by-one in this manner, and the generalization error can be calculated in a fraction of the time it would take to retrain a single new network. Once a limit on the number of irrelevant nodes is determined, a new, smaller, network can be retrained. Let us now consider the application of this method.

For simplicity, consider a feed-forward network with one input layer, one hidden layer and an output layer. The weighted sum input (Equation 3.15), of the p^{th} input pattern, to the i^{th}, output node is

$$i_i^p = \sum_j w_{ij} h_j^p + \theta_i , \qquad (3.52)$$

where h_j^p is the activation value of the $j^{th}, j = 1 \ldots N$, hidden layer node. In vector notation,

$$i_i^p = w_i^T \mathbf{h}^p + \theta_i \ . \tag{3.53}$$

This linear function is particularly suited for the PCA analysis discussed above. While standard PCA is useful in data compression, that is mapping $\Re^N \rightarrow \Re^M$ with $M \leq N$ where the smaller M-dimensional space contains the majority of the discriminatory information as described by Equation 3.50, we are not specifically interested in reducing the input dimensionality. Rather, we can apply the method in an altered form to study the effect of removing discriminatory information contained in the data. If we consider the N-dimensional input space to be the outputs of the N hidden layer nodes, we would like to: calculate the class-discriminating principal components (by construction, there will be N principal components), rotate the input space to the PCA space, project out the components corresponding the M lowest eigenvalues, and rotate back to the original N-dimensional input space. Using the *ansatz* that a network defined with $(N - M)$ hidden units would reconstruct the remaining $(N - M)$ principal components, we calculate the generalization error using these redefined inputs, and if there is no significant effect we say that we have pruned the hidden layer.

More specifically, using the algorithm defined in the previous section, if $\bar{\mathbf{h}}$ is the average vector of the outputs of the hidden layer, $A \in \Re_{N \times N}$ is the related PCA rotation, $\mathcal{M} \in \Re_{N \times N}$ is a matrix which projects out the last $N - M$ eigenvectors,

$$\mathcal{M} = \begin{pmatrix} 1_{M \times M} & 0 \\ 0 & 0 \end{pmatrix} \ . \tag{3.54}$$

We define the projected hidden layer output to be

$$\mathbf{h}^p \mapsto AM \left[(A^T(\mathbf{h}^p - \bar{\mathbf{h}})) \right] + \bar{\mathbf{h}} \ . \tag{3.55}$$

As it stands, this transformation must be applied to each input pattern and this would be very computationally intensive. However, once calculated, the PCA transformation is constant so we can redefine the network instead of the inputs.

Applying Equation 3.55 to Equation 3.53, we obtain,

$$\begin{aligned} i_i^p &= w_i^T AM \left[(A^T(\mathbf{h}^p - \bar{\mathbf{h}})) + \bar{\mathbf{h}} \right] + \theta_i \\ &= w_i^T (AMA^T)\mathbf{h}^p + w_i^T (1 - AMA^T)\bar{\mathbf{h}} + \theta_i \\ &= (AMA^T w_i)^T \mathbf{h}^p + \left(w_i^T (1 - AMA^T)\bar{\mathbf{h}} + \theta_i \right) \\ &= \tilde{w}_i^T \mathbf{h}^p + \tilde{\theta}_i \ , \end{aligned} \tag{3.56}$$

with,

$$\tilde{w}_i \equiv AMA^T w_i \tag{3.57}$$

$$\tilde{\theta}_i \equiv \theta_i + w_i^T (1 - AMA^T)\bar{\mathbf{h}} \ . \tag{3.58}$$

Checking two extreme cases, if $\mathcal{M} = 1_{N \times N}$, then $\tilde{w}_i^T = w_i^T$ and $\tilde{\theta}_i = \theta_i$, as we would expect since no information has been removed. If $\mathcal{M} = 0_{N \times N}$, then $\tilde{w}_i^T = 0$ and $\tilde{\theta}_i = \theta_i + w_i^T \bar{\mathbf{h}}$, which yields $i_i^p = w_i^T \bar{\mathbf{h}} + \theta_i$. In this case, once all discriminatory information is removed, the only thing remaining is the unit threshold θ_i and the mean spectrum $\bar{\mathbf{h}}$; no class discrimination is possible.

In the intermediate case, when some number $(N - M)$ eigenvectors are projected out, $\mathcal{M} = 1_{M \times M} \oplus 0_{(N-M) \times (N-M)}, M < N$, there may be a node which has most of its relevant information removed, and it would approach the case where its output is equivalent to its mean value for all input types, thus simulating the case where that node provides no discriminating power to the network.

Alogrithm

All neural networks studied in this thesis can be classified as a simple feed-forward, multi-layer network with a input layer; hidden layer(s); and output layer architecture.

The following algorithm is used to calculate the effective number of hidden units needed:

- In supervised learning, we present the network with a set of training events. The N output values of the hidden layer units for each pattern p are represented by $\mathbf{h}^p$.
- We then determine the mean spectrum of the output values of the hidden layer units, represented by the vector $\bar{\mathbf{h}}$.
- Using $\bar{\mathbf{h}}$ to mean-shift the outputs, the difference vector covariance matrix C is calculated.
- The eigenvectors $\mathbf{a}_i$ and related eigenvalues λ_i of the covariance matrix C are then calculated. As explained above, the eigenvectors corresponding to the largest eigenvalues are those which contain the greatest amount of discriminatory information.
- The eigenvectors $\mathbf{a}_i$, ordered by deceasing eigenvalues λ_i, are the column vectors of the rotation matrix A which diagonalizes C, such that $\mathrm{Tr}[ACA^T] = \sum_i \lambda_i$.
- The $(N - M)$ least significant eigenvectors are projected out of the rotated input space via $\mathcal{M} = 1_{M\times M} \oplus 0_{(N-M)\times(N-M)}, M = 0 \ldots N$.
- A, $\bar{\mathbf{h}}$ and $\mathcal{M}$ are used to recalculate the internal weights and thresholds, $\tilde{w}$ and $\tilde{\theta}$, connecting the hidden layer to the output units.
- For each projection, the generalization error of the network is determined and compared to the original network error (corresponding to $\mathcal{M} = 1$).
- The network is considered maximally pruned when further projections significantly begin to degrade the generalization error of the network.

This algorithm is diagrammatically depicted in Figure 3.9.

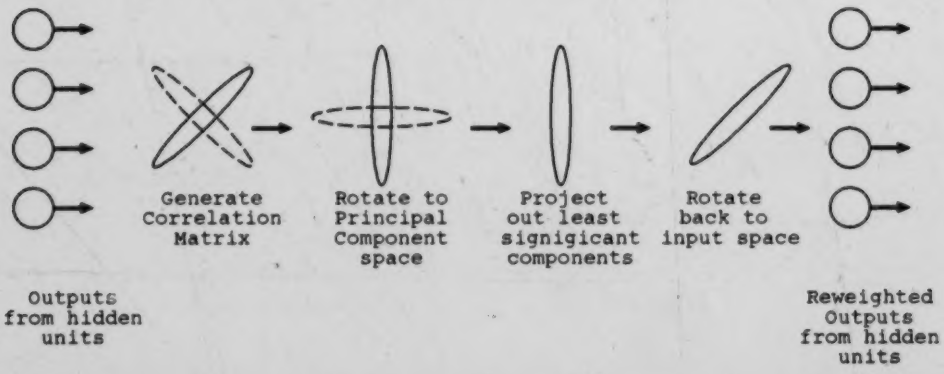

Generate Correlation Matrix
Rotate to Principal Component space
Project out least signigicant components
Rotate back to input space
Outputs from hidden units
Reweighted Outputs from hidden units

Figure 3.9: A diagrammatic respresentation of the Pricipal Component Pruning algorithm. The outputs from the hidden units are measured and used to calculate their correlation matrix, which in turn defines the Principal Component rotation. Rotating to the Principal Component space, the direction with the least significant eigenvalues are projected out. The surviving eigenvectors are rotated back to the input space and used to re-weight the weights and thresholds connecting the hidden layer to the output layer.

Application

To test this algorithm, a simple test case was considered. Two data classes, $C = X$ and Y, were simulated by generating two data sets, each consisting of five random parameters, gaussian distributed with covariance matrix V_C and mean values zero. The five parameters in each sample were related by an arbitrary covariance matrix, V_C, with $V_X \neq V_Y$. In Table 3.1 we see the results of the Principal Component Discrimination Analysis applied to these sample data points. We can see that three principal components contain $\sim 95\%$ of the information needed to discriminate between the two test classes. We then train a small neural network with five inputs, 20 units in the hidden layer and a single output unit. The output unit is trained to be 1 for events from class X and 0 for events from class Y.

Eigenvalue	λ	% of Trace	cum. % of Trace
1	29.688	39.784	39.784
2	25.656	34.381	74.164
3	16.989	22.767	96.932
4	1.320	1.769	98.700
5	0.970	1.300	100.000

Table 3.1: Principal Component Discrimination Analysis applied to the raw inputs of the example distributions.

Eigenvalue	λ	% of Trace	cum. % of Trace
1	1.7695	31.1446	31.1446
2	1.2863	22.6401	53.7846
3	1.0635	18.7185	72.5031
4	0.4620	8.1315	80.6346
5	0.3804	6.6956	87.3302
6	0.3342	5.8816	93.2119
7	0.1362	2.3971	95.6090
8	0.0662	1.1653	96.7743
9	0.0603	1.0621	97.8365
10	0.0413	0.7269	98.5634
11	0.0313	0.5513	99.1147
12	0.0207	0.3643	99.4790
13	0.0111	0.1960	99.6750
14	0.0060	0.1049	99.7799
15	0.0055	0.0976	99.8776
16	0.0025	0.0446	99.9222
17	0.0022	0.0389	99.9611
18	0.0010	0.0180	99.9791
19	0.0008	0.0141	99.9932
20	0.0004	0.0068	100.0000

Table 3.2: Principal Component Analysis applied to the outputs of the twenty units in the hidden layer. $\sim 95\%$ of the discrimination power is contained in the first 6 eigenvectors.

Table 3.2 presents the Principal Component Analysis applied to the outputs of the ten units in the hidden layer. It can be seen that the first ten eigenvectors contain $> 98\%$ of the discrimination power between the two classes, with the last ten eigenvectors contributing less than 2%. The next

Eigenvectors Used	% Change in Generalised Error
20	0.000
19	0.001
18	0.001
17	0.004
16	0.004
15	0.006
14	0.056
13	0.022
12	0.021
11	0.266
10	0.188
9	0.302
8	0.353
7	0.335
6	3.361
5	8.026
4	173.532
3	209.583
2	218.363
1	230.744

Table 3.3: Principal Component Pruning applied to the outputs of the twenty units in the hidden layer. The lefthand column indicates the number of ordered principal component eigenvectors remaining after the projection was applied. The righthand column shows the effect of the projection on the generalised error, calculated by applying the network to a data set not used in the training phase. We see that projecting out the ten least significant eigenvectors has little effect of the generalised error of the network, indicating that a network trained with ten hidden units should be sufficient to efficiently classifiy the data classes.

step in the algorithm would be to project out the least significant eigenvectors by adjusting the internal weights and thresholds *in situ*, and then calculating the generalized error of each new network.

In Table 3.3 we see the effect on the Principal Component Pruning method. The left hand column indicates the number of ordered principal component eigenvectors remaining after the projection was applied. The right hand column shows the effect of the projection on the generalized error, calculated by applying the network to a data set not used in the training phase. We see that projecting out the ten least significant eigenvectors has little effect of the generalized error of the network, $\approx 0.2\%$, indicating that a network trained with ten hidden units should be sufficient to classify the data classes.

This is most easily checked by comparing the outputs of the different networks directly, each identical except for the number of units in the hidden layer. Figure 3.10 presents the classification efficiencies, between classes X and Y, for these networks. In agreement with the results of the Principal Component Pruning method, presented in Table 3.3, there is little classification difference between the twenty and ten hidden unit networks. We begin to see a loss in classification efficiency for the signal class, at a constant background efficiency, for a five hidden unit network, again in agreement with prediction. Significant loss in classification power is observed for networks with still fewer hidden units.

3.5 Systematic Error Checks with Neural Networks

3.5.1 Transformation of Monte Carlo Distributions

In order to study the systematic uncertainties that arise from the imperfect modeling of the variables used by the ANN, a mechanism has been developed to transform the monte carlo values of a given variable (the *original* distribution) to yield a new value that is consistent with the distribution observed in the data (the *target* distribution). As the method is quite general, we will choose to refer to the distributions as *original* and *target*, and later apply the results to some actual monte carlo and data distributions.

The transformation of a continuous [1] variable, $x_o \in [a, b]$, which yields values consistent with the target distribution, $x_t \in [a', b']$. This transformation can be achieved by calculating the fractional probability integral.

Let the original continuous distribution be represented by $f_o(x)$, defined on $x \in [a, b]$. This curve is typically obtained by a fit to a histogram and, once properly normalized, lends itself to the interpretation of a probability density. Let us now define $F_o(x_o)$ as the normalized fractional integral

$$F_o(x) = \frac{\int_a^x f_o(z)dz}{\int_a^b f_o(z)dz} .$$

Similarly, for the target distribution we define $f_t(x)$ and $F_t(x)$. For any original value, x_o we may associate $y_o = F_o(x_o)$. Supposing we can invert $F_t(x)$, the desired transformation may be written

$$x_t = F_t^{-1}(y)\big|_{y=F_o(x_o)}$$

. In essence, this is just an application of the standard [80] method for generating a known distribution from a uniform distribution. The key element here is that we generate the required uniform distribution from the original distribution.

After obtaining the fractional probability integrals for both the original and target distributions, the transformation can be carried out in the following manner:

[1] The transformation of discrete variables is not consistent with this method. If a monte carlo event is generated with a single track, it would be difficult to accept a transformation which resulted in two, however, if that same event were generated with, for example, $\sqrt{s'}$, then a transformation to $\sqrt{s'} + \epsilon$ could be desirable to better agree with the observed distribution of $\sqrt{s'}$ in the data.

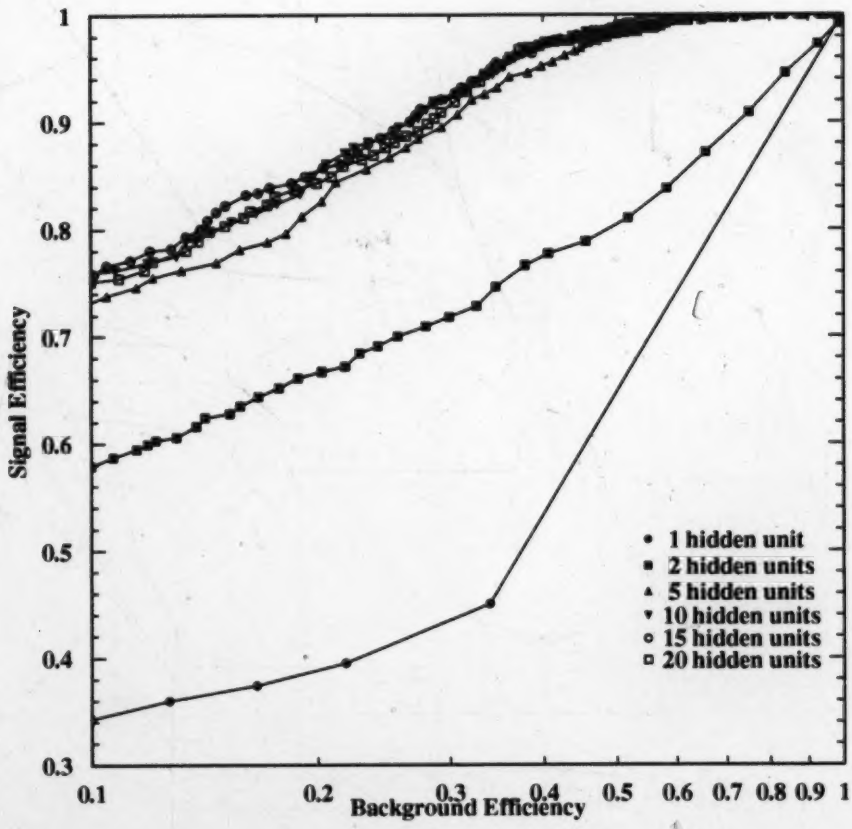


Figure 3.10: Presented are the classification efficiencies, between signal class X and background class Y, of a number of networks, differing only in the number of units in the hidden layer. In agreement with the results of the Principal Component Pruning method, presented in Table 3.3, there is little classification difference between the twenty and ten hidden unit networks. We begin to see a loss in classification efficiency for the signal class, at a constant background efficiency, for a five hidden unit network, again in agreement with prediction. Significant loss in classification power is observed for networks with fewer still hidden units.

- Obtain the value of the variable of interest for a particular event.
- Determine the value of the fractional probability integral obtained from the original distribution for this value of the variable.
- Taking this value for the fractional probability integral, invert target fractional integral evaluated at the given ordinate.
- The abscissa of the inversion is the value that should be used for the transformed variable

Implementation

The easiest implementation of the code would be to use the binned histograms to quickly calculate approximate fractional integrals.

The original data point, x_o, would be matched to a bin, i, in the original distribution. The fractional integral would be calculated relative to the centre of that bin, x_i, yielding a value y_o. One could then step through the target fractional integral and determine the bin j, such that $y_{j-1} \leq y_o < y_{j+1}$ and the bin centre, x_j would be the transformed value, x_t.

This suffers from a number of problems. First, we can clearly see that we are in fact mapping a continuous distribution to a discrete number of bins centres in the target distribution. Second, if two bins in the target distribution happen to have zero entries, the fractional integral will not change over this range and the inversion will not be analytic.

Evaluation of the Fractional Integral

The original and target distributions are first read in, and filled, to histograms of constant size. The choice of binning here is typically determined by the user as algorithmic methods are not in common use. One should aim to maximize the number of bins, N_{bins}, to avoid unnecessary loss of information, while ensuring there is a statistically adequate number of events in each bin.

The range of the distribution is also worthy of consideration. One would usually choose to minimize this range, since the inclusion of regions where the distributions are unpopulated will make it difficult to choose a binning which maintains a reasonable number of events in each bin. The effect of not including unpopulated regions will hopefully be minimal because, by definition, they do not contain a large fraction of the events.

The distributions are then spline fit using the `NAGLIB` [81] routine `E02BEF`, a least-squares cubic spline curve fit with automatic knot placement. This spline returns a functional representation of the distribution which we can later evaluate using the associated routine `E02BBF`. This fitting plays two roles in this analysis; not only does it allow the accurate evaluation of the fractional integrals, but also smooths the data, which is a desirable property when dealing with target distributions of real data where there may be significant statistical fluctuations.

The `NAGLIB` routine `E02BBF` will not allow the user to query the spline outside its defined range. To avoid potential problems, the derivatives at the end points were evaluated using the `NAGLIB` routine `E02BCF` allowing for a linear extrapolation of the fit for the over and underflow points. Extrapolation in general is undesirable. However, we expect it to be a reasonable mapping within a "small" region outside the defined range. If we choose this small region, Δ, to be on the order of the mapping difference at the end points of the spline, $\Delta = |x_t^{\text{End}} - x_o^{\text{End}}|$, we can make the final transformation well defined and accurate over the whole real line by smoothly deforming the extrapolation to a 1:1 mapping over this reasonable range Δ.

The spline fit is relatively automated and requires only one user input, $\mathcal{S}$, which specifies the desired χ^2 between spline fit and the original distribution. If one assumes the spline fit is a representation of the parent distribution from which our finite sample distribution was generated, then any discrepancy will be due to statistical fluctuations, and so choosing, $\mathcal{S} = \mathcal{S}_{\text{ndf}}(N_{\text{bins}} - 1)$

will make S an estimator of χ^2 with a $\chi^2/\text{ndf} = \mathcal{S}_{\text{ndf}}$ [82]. Thus, choosing $\mathcal{S}_{\text{ndf}} = 1.0$ will yield a spline fit consistent with the variance due to statistical fluctuations [2].

By choosing $\mathcal{S}_{ndf} < 1.0$, the fit will follow the fluctuations in the original distribution more closely, thus we could simulate the effect of structure in the underlying distribution which may not be due to statistical fluctuations.

Using `E02BBF` to query the spline fit, we numerically calculate the fractional integral by calling the `CERNLIB` routine `DGAUSS`. As previously, mentioned we are also required to invert the fractional integral of the target distribution, $F_t(x)$. Given the observation x_o, the desired transformation is achieved by numerically calculating the zero of the function $F_t(x) - F_o(x_o)$ using the `CERNLIB` routine `DZERO`. The accuracy of the determined zero is set by the user and should be at least an order of magnitude less than the accuracy of the evaluation of the fractional integral to ensure reasonable results.

Ensuring a Monotonically Increasing Fractional Integral

The method of inversion by finding the zero of $F_t(x) - F_o(x_o)$ is an obvious choice but needs some attention. If there exists a finite range of the target distribution where the spline (distribution) is zero, then the inversion will be ill-defined. By adding a constant bias to the distribution, we ensure a monotonically increasing fractional integral and thus a well defined inversion. This bias is added to both the numerator and denominator integrations in the definition of the fractional integral so will not affect the results. However, the magnitude of the bias is important; if it is smaller than the accuracy of the fractional integral evaluation its effect will not be noticed and the inversion may become ill-defined, and if its magnitude is much larger than the distribution, then the distribution itself will become lost in the integration and the fractional integral will approach a straight line with unit slope. In this analysis, we chose to set the bias an order of magnitude greater than the accuracy of the fractional integral evaluation.

Speed and Accuracy

The method of numerically evaluating the zero of the function $F_t(x) - F_o(x_o)$ requires the fractional integral to be evaluated many times for each inversion and soon becomes quite computationally intensive for any substantial number of inversions. The accuracy of the method, however, easily can be many orders of magnitude smaller than the bin size of the distributions, which we will assume to be the minimum size for the given statistics and experimental precision of the available data.

Optimally, we need only invert specific points in the distribution and then use these points for interpolating the transformation of the original distribution. As long as we choose a sufficiently large number of points to calculate the inversion, the interpolation using the `CERNLIB` routine `DIVDIF` will allow for the required transformation accuracy. This interpolation can then be incorporated into a stand-alone routine to be used in the users main analysis.

Application

An example which illustrates this procedure is shown in Figure 3.11.

Plot '(a)' in the top left of the figure shows the original distribution, the one which we wish to map to the target distribution. In this case it is a Gaussian distribution $f(x) = A\exp\left(\frac{(x-\mu)^2}{\sigma^2}\right)$ with $\mu = 1.0, \sigma = 0.5$ and $A = 1$ was used to generate $10K$ events. The plot itself has been scaled to agree with the integral of the target distribution. The histogram is fit with a gaussian and we can see the fitted values for μ, σ as are expected.

Plot '(b)' in the top right of the figure shows the target distribution. In this case, the sum of two Gaussians, $f(x) = A_1\exp\left(\frac{(x-\mu_1)^2}{\sigma_1^2}\right) + A_2\exp\left(\frac{(x-\mu_2)^2}{\sigma_2^2}\right)$, with $A_1/A_2 = 2$, $\mu_1 = 2, \mu_2 = 4$

[2] Choosing $\mathcal{S}_{ndf} = 0.0$ will result in `E02BBF` returning an interpolating spline fit

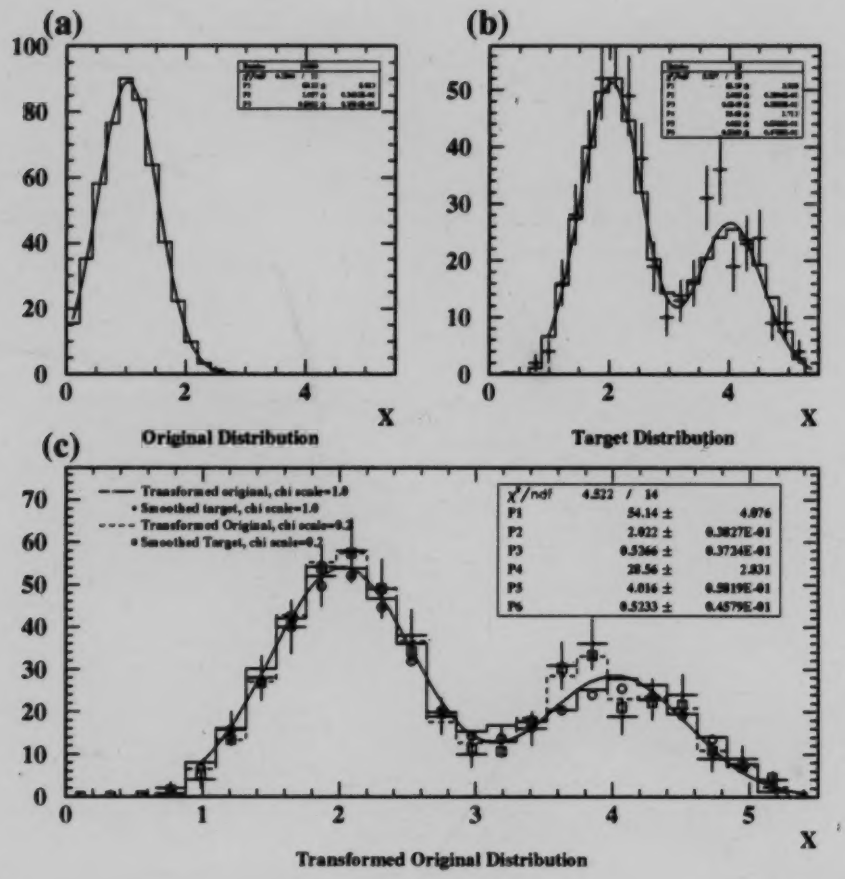


Figure 3.11: Plot (a) shows the original gaussian distribution. Plot (b) shows the target distribution as the sum of two gaussians, with the smoothed spline fit overlayed. Plot (c) presents the results of the method. The crossed points give the raw, unsmoothed target distribution and the open circles the smoothed result with $\mathcal{S}_{ndf} = 1.0$. The solid histogram is the transform of the original distribution and a fit, presented by a solid line, to this transformed histogram shows good agreement with the target distribution.

and $\sigma_1 = \sigma_2 = 0.5$, was used to generate a sample of 500 events, represented by the crossed data points. Infrequently would it be the case that a simulated distribution would be so different from the observed data distribution, but we will show that the method can easily handle this gross example.

The statistical fluctuations in the target distribution are representative of analysis situations where a small data sample is available for comparison to a much larger monte carlo set. As mentioned earlier, to handle this, the distributions is smoothed using a spline fit. This smoothed spline fit distribution, using $\mathcal{S}_{\rm ndf} = 1.0$, is represented by the solid histogram. A fit of the sum of two gaussians to this histogram is shown by the solid curve and fit results presented show good agreement with their expected values.

Plot '(c)' at the bottom of the figure presents the distributions of interest for this method. The crossed points give the raw, unsmoothed target distribution, and the open circles the smoothed target with $\mathcal{S}_{\rm ndf} = 1.0$. The solid histogram is the transformation of the original distribution and a fit, presented by a solid line, to this transformed histogram shows good agreement with the target distribution. Before proceeding we emphasize that the distribution in plot '(a)' has been ***smoothly*** transformed to plot '(b)'. The agreement of the transformed distribution (solid histogram) to the smoothed target (open circles) appears to be quite accurate.

Changing the value of $\mathcal{S}_{\rm ndf}$ is expected to give the user the ability to follow the fluctuations in the target distribution as closely as they like. In plot '(c)' we show the smoothed fit to the crossed points for a $\mathcal{S}_{\rm ndf} = 0.2$ as open squares. As expected, the fluctuations in the raw target distribution are followed much more closely, and the dashed histogram shows the transformation of the original distribution is mapped quite well again.

3.5.2 ANN Systematic Error

So far, we have only been concerned with mapping an arbitrary original distribution to an arbitrary target distribution given some accepted accuracy. If we re-interpret this method, however, we will find that this may be used in systematic error studies.

All systematic error sources in an analysis will contribute to the variations observed between the associated data and monte carlo distributions. In a standard systematic error analysis one varies the relevant parameters (or equivalently the cuts on those parameters) assuming, or ensuring, no correlation between different systematic error contributions, and the final systematic error is then typically a sum in quadrature of all those contributing sources.

One important feature of this transformation method, unlike standard re-weighting techniques, is that correlations between variables are maintained. The degree of correlation may differ from the original, but since the transformation performs a one-to-one mapping from the original to transformed value, correlations are not destroyed.

It would seem, then, that a transformation of all parameters relevant to an analysis can be used to determine the total systematic error in a way which accounts for all the inherit sources of systematic error, both known and ***unknown***, as well as dealing with their correlations in a consistent manner.

One caveat is that if the data distribution is consistent with a signal plus background hypothesis, rather than a background only hypothesis, a background subtraction method should be used first so that the transformation of the monte carlo background will not include artifacts of the signal events. This obviously adds a further level of complexity, and in this case, the ability to make an interpretation of the method as an estimator of the total systematic error is less clear.

Chapter 4

The Search for $e^+e^- \to Z^0 \to h^0A^0 \to A^0A^0A^0 \to 3(b\bar{b})$ at $\sqrt{s} = 183\,\mathrm{GeV}$

4.1 Introduction

The results presented in this chapter detail the search for $h^0\,A^0 \to A^0\,A^0\,A^0 \to 3(\,b\bar{b}\,)$ in the region of the $(\,m_{h^0}, m_{A^0}\,)$ mass plane satisfying $2m_{A^0} \leq m_{h^0}$, parts of which were left unexcluded by the previous $\sqrt{s} = 172\,\mathrm{GeV}$, as discussed in Section 1.2.2. Due to the presence of six b-quarks the events are characterized by a large number of jets and a large charged track multiplicity. To reduce backgrounds, b-tagging plays a crucial role. At 183 GeV, backgrounds from $(Z/\gamma)^* \to b\bar{b}\,g(\gamma)$ with hard gluon emission and four-fermion processes contribute approximately equally. Backgrounds from two-photon processes are reduced to a negligible level by the event selection. The analysis presented here is based on data collected by the OPAL detector during the 1997-1998 run, representing a luminosity of $\mathcal{L} = 54.5 \pm 0.2\ \mathrm{pb}^{-1}$. Selection of candidates is done through a neural network analysis which combines kinematic and topological variables with heavy flavour tagging.

4.2 Data and Monte Carlo

To simulate the signal, the process $e^+e^- \to h^0\,A^0 \to A^0\,A^0\,A^0 \to 3(\,b\bar{b}\,)$ was generated, with various m_{h^0}, m_{A^0}, at 184 GeV using `HZHA` [29] with the Standard Model parameters $\alpha_s = .118, m_{top} = 175\,\mathrm{GeV}, m_{Z^0} = 91.189\,\mathrm{GeV}, \Gamma_Z = 2.497\,\mathrm{GeV}, G_F = 1.16639 \times 10^{-5}\,\mathrm{GeV}^{-2}$, and full simulation of initial state radiation. The main contributions to the total background come from those processes which have high charged track multiplicity and which can mimic multi-jet final states. The most important backgrounds contributing to this analysis were composed of the two fermion processes (with initial state radiation), $(Z/\gamma)^* \to q\bar{q}$, generated by `PYTHIA` [83], and the four fermion final states from W^+W^- and $Z^0Z^{0(*)}$ generated using `GRC4F` [84]. For systematic checks, described more fully below, an alternative Standard Model background sample was generated using `HERWIG` [85] to simulate the $(Z/\gamma)^* \to q\bar{q}$ contribution and `PYTHIA` for the remaining allowed W^+W^- and $Z^0Z^{0(*)}$ channels.

Events were then processed through a full simulation of the OPAL detector [86] using the `GOR019` and `GOR021` descriptions of the detector layout with the `OPAL97` constants and calibrations.

The first available Monte Carlo samples had been generated at $\sqrt{s} = 184$ GeV. The data runs, however, were found to have energies closer to $\sqrt{s} = 183$ GeV, so new Monte Carlo samples were generated due to threshold effects involved with some of the background channels, namely W^+W^-

production. The main background sample used in the this analysis was generated at the correct $\sqrt{s} = 183\,\text{GeV}$. However, the available $\sqrt{s} = 184\,\text{GeV}$ Monte Carlo was still useful, especially for the two fermion processes which were not sensitive to threshold effects. Some of this 184 GeV Monte Carlo was included in the alternative background samples used in the systematic studies. The signal and Standard Model background samples used to optimize the selection cuts are listed in Table 4.1

Process	Run	Number of Events	[pb]
Data, Pre CV Incident			
Data	Period 84, Pass 5	716	6.25
Data	Period 85, Pass 5	789	7.00
Data	Period 86, Pass 5	2593	22.84
Data	Period 88, Pass 5	162	1.47
Data, Post CV Incident			
Data	Period 88, Pass 4	1488	13.40
Data	Period 89, Pass 4	383	3.53
HZHA generated signal, $hA \to 3(b\bar{b})$			
$m_{h^0} = 60,\ m_{A^0} = 30$	...L20010/P184HA/R6725/P0008	0.5K	
$m_{h^0} = 60,\ m_{A^0} = 30$	...L20010/P184HA/R7048	5K	
$m_{h^0} = 80,\ m_{A^0} = 30$	...L20010/P184HA/R6725/P0009	0.5K	
$m_{h^0} = 80,\ m_{A^0} = 30$	...L20010/P184HA/R7049	5K	
$m_{h^0} = 80,\ m_{A^0} = 40$	...L20010/P184HA/R6725/P0010	0.5K	
Main PYTHIA, GRC4F background			
$Z^0(\gamma) \to q\bar{q}$	...L20011/P183QQ/R5050	375K	107.43
$e^+e^- \to q\bar{q}q\bar{q}$	...L20011/GRC4F/R7051	30K	7.86
$e^+e^- \to q\bar{q}\ell^+\ell^-$	...L20011/GRC4F/R7050	40.6K	8.11
$e^+e^- \to q\bar{q}e^+e^-$	...L20011/GRC4F/R7055	100K	26.73
Alternative HERWIG, PYTHIA background			
$Z^0(\gamma) \to q\bar{q}$	...L20010/ZZHW59/R1124	150K	101.42
$W^+W^- \to$ all	...L20011/GRC4F/R7337	25K	15.54
$Z^0Z^{0(*)} \to$ all	...L20011/P183ZZ/R7338	5K	0.61

Table 4.1: Monte Carlo and OPAL Data used in the 183 GeV analysis. $e^+e^- \to q\bar{q}\ell^+\ell^-$ samples do not include $e^+e^- \to q\bar{q}e^+e^-$. The right hand column reports the luminosity represented by the data samples, or the cross-section of the background process.

During the 1997 data taking period, a set of three wires in sector 26 of CV stereo broke at event 12773 in run 8679 (period 88) [87]. CV axial and stereo were down for a day, and the axial portion was ramped up by run 8692. Two thirds of the stereo section was ramped up starting with run 8816. Runs 8679 (starting at event 12773) –8691 have CV detector status of 1 (see Table 4.2), and runs 8692 onwards have CV detector status of 2. New calibrations of the stereo layer were necessary as the field configurations had changed. Therefore, in order to monitor data quality, the data was considered as pre and post CV incident.

4.3 Preselection

The full data set collected by OPAL during the 1997 data taking period amounted to more than 75000 multi-hadronic events. Due to this large number, it was impractical to analyse and import all the available data. Rather, a cascading set of loose selection cuts were imposed to filter out interesting events for further analysis. The cuts were selected with the aim to obtain a final sample of signal candidate events which had as little contamination from other event types as possible. This *preselection* is a multi-step process.

When OPAL data is recorded, a number of easily calculable quantities, such as the event's energy and number of charged tracks, are stored with each event. Selections based on these variables require no additional computing, making them very fast, while significantly reducing the number of events

to be analysed. Further computing intensive analysis was then only applied to those events passing the initial selections.

The preselection cuts used in this analysis will be described in the following sections. At times, a detailed description or motivation for a particular variable will be given, as well quantities not directly related to the preselection, but relevant to later discussions, may be described when they are logically related to a preselection variable. Because of these additional explanations, the preselection cuts will be summarized in Section 4.3.8.

4.3.1 Subdetector Status

Each subdetector provides a status number for every event recorded. These status numbers, and their meanings, are listed in Table 4.2. To begin the analysis, the status of the relevant subdetectors were checked to ensure they were working well. For this analysis, the following subdetectors were required to have good status flags: `SI`, `CJ`, `EB`, `EE`, and `HT` [1], as well as `SW` (to ensure an accurate luminosity calculation). The CV subdetector was required to have a good status flag before the CV stereo incident and a status of ≥ 2 after.

Status	Interpretation
0	Detector is dead
1	Detector is unreliable
2	Detector has small problems
3	Detector is reliable

Table 4.2: Detector status values and their interpretation.

Using only the detector status selection, the total luminosity for the data set analysed was calculated, using the OPAL analysis routine `ROCROS` [88], and found to be $\mathcal{L} = 54.5 \pm 0.2$ pb^{-1}.

4.3.2 Hadronic Event Selection

The OPAL selection of hadronic events uses a combination of tracking and calorimetric requirements. In order to be used in forming this selection, the reconstructed tracks and clusters must pass certain quality requirements listed in Table 4.3, where $E_{\rm Vis} = \sum_{\rm N_{Clus}} E_{\rm Raw}$ is the visible energy in the event, and $E_{\rm Bal} = \sum_{\rm N_{Clus}} E_{\rm Raw} \cos\theta$ is the balanced energy, the sums being taken over the number of good clusters, $\rm N_{Clus}$. The first set of conditions defines the standard LEP2 Multi-Hadron (L2MH) [89] event selection. Additionally, due to the six jet nature of the signal, a higher number of tracks are expected. So, to reject Standard Model background, mainly $e^+e^- \to (Z/\gamma)^* \to q\bar{q}(g)$, the minimum number of good tracks was increased to 25.

4.3.3 Event Reconstruction

Once these basic preselections are made, the more computing intensive phase of the analysis begins. The analysis variables were generated using Higgs group analysis package, `DH` [90], along with additional, signal specific, code added by the author.

Track and Cluster Quality

Tracks, electromagnetic clusters and hadronic clusters are identified using the selection criteria in Table 4.4. These are similar to those discussed in [91], with the exception that the default value for the maximum allowed track momenta was raised from 60 GeV $\to$ 120 GeV to reflect the higher

[1] The Hadron Calorimeter Towers data acquisition system

LEP2 Multi-Hadron (L2MH) selection		
Tracks	Number of Central Detector hits, N_{CD}	≥ 20
	$\lvert d_0 \rvert$	$\leq 2.0\,\text{cm}$
	$\lvert z_0 \rvert$	$\leq 40.0\,\text{cm}$
	$\text{radius}_{1^{st}\text{hit}}$	$\leq 60.0\,\text{cm}$
	p_{Trk}^{T}	$\leq 0.050\,\text{GeV}$
	$\lvert \cos\theta_{Trk} \rvert$	≤ 0.995
	$\chi^2_{r-\phi}$	≤ 999
	χ^2_{s-z}	≤ 999
Clusters	E_{Raw}	$\geq 0.100\,\text{GeV}$ in the Barrel $\geq 0.200\,\text{GeV}$ in the Endcaps
	N_{Blocks}	≥ 1 in the Barrel ≥ 2 in the Endcaps
Event	$R_{\text{Vis}} = E_{\text{Vis}}/2\sqrt{s}$	≥ 0.14
	$\lvert R_{\text{Bal}} \rvert = \lvert E_{\text{Bal}}/E_{\text{Raw}} \rvert$	≤ 0.75
	$N_{\text{Good Clusters}}$	≥ 7
	$N_{\text{Good Tracks}}$	≥ 5
Additional Constraints		
Event	$N_{\text{Good Tracks}}$	≥ 25

Table 4.3: Tracks and cluster quality cuts in the LEP2 Multi-Hadron (L2MH) selection. These selection criteria were developed to efficiently select samples of multi-hadronic events at LEP2 energies. The additional constraint on the number of good tracks was imposed to increase the rejection of $e^+e^- \rightarrow (Z/\gamma)^* \rightarrow q\bar{q}$ events.

$\sqrt{s}$ at LEP2 . Information from the Silicon detector is used to improved track reconstruction using the OPAL Silicon detector track matching routine, ODSIZL [87].

Analysis Level Event Quality		
Tracks	p_{Trk}^{T}	$\geq 0.100\,\text{GeV}$
	p_{Trk}	$\leq 120\,\text{GeV}$
	Number of CJ hits	≥ 20
	$\lvert d_0 \rvert$	$\leq 2.5\,\text{cm}$
	$\lvert z_0 \rvert$	$\leq 50.0\,\text{cm}$
	χ^2_{s-z}	≤ 100
	$\lvert \tan\lambda \rvert$	≤ 3.73
Clusters	E_{Raw}	$\geq 0.170\,\text{GeV}$ in the Barrel
		$\geq 0.250\,\text{GeV}$ in the Endcaps
	N_{Blocks}	≥ 1 in the Barrel
		≥ 1 in the Endcaps
Towers	E_{Tower}	$\geq 0.250\,\text{GeV}$

Table 4.4: Analysis level event quality selections.

As each track must be well measured to determine its energy and momentum, only those tracks which satisfy the quality selections were used in this analysis. A minimum number of Jet Chamber (CJ) hits, N_{CJ}, on each track is demanded and the track must have: a minimum angle (corresponding to a maximum $|\cos\theta_{\text{Trk}}|$); a minimum momentum p_{Trk}; and a minimum transverse momentum, with respect to the beam, p_{Trk}^{T}. In addition to being well measured, each track should appear to originate from the primary interaction point, thus, cuts are made on z_0, the apparent z-coordinate from which the track originated, and d_0, the distance at the point of closest approach of the extrapolated track to the interaction point. Those which do not appear to come from the interaction point may be artifacts due to jet chamber sparking, beam wall and beam gas interactions, or backscatter in the solenoid from particles which have already left the jet chamber.

Clusters used in this analysis should not be due to either poorly calibrated blocks or noisy electronics. Each cluster is compared to a list of noisy blocks and, if a match is made, it is dropped from further analysis. The energy of a cluster must then be corrected to account for the energy deposited in the material in front of the lead-glass. This is a large correction for clusters with a small deposited energy, E_{Raw}, so a minimum cut is made on E_{Raw}. For a cluster in an electromagnetic endcap, which has its blocks mounted parallel to the beam axis, a minimum number of blocks, N_{Blocks}, was required.

4.3.4 Track-Cluster Matching

In the hadronic final state, charged particles carry about 2/3 the total energy. The energy of these charged tracks is measured twice: once in the central tracking system, using their curvature and $\mathrm{d}E/\mathrm{d}x$ information, and a second time by direct measurement in the electromagnetic and hadronic calorimetry. The neutral hadronic energy, however, can only be measured with calorimetry.

For the OPAL detector, it is often the case that momenta and angular resolutions for charged particles are measured much more accurately with the central tracking system than with the calorimeters. So, we would like to define the event energy as the sum of the charged track energies, measured with the central detector, with the neutral hadronic energy measured in the calorimetry. To accomplish this, one needs to avoid the double-counting of energy from a good quality track leaving a good quality cluster.

Naively, we could hope to remove this double counting by using only those clusters not associated with any tracks. However, this would require that the clusters do not overlap which is incorrect.

Rather, an algorithmic approach, using the MT [92] package, was used to subtract off the energy of well measured tracks from their associated clusters.

4.3.5 Jet Reconstruction

One of the most striking, and in fact identifying, features of an hadronic event is the presence of *jets*.

The lowest order process for producing hadrons in e^+e^- colliders is $e^+e^- \to q\bar{q}$, where the primary quarks fragment into hadrons. The essential assumption of the parton model, leading to the expectation of jets, is that the transverse momenta of the quark fragmentation products arise mainly from soft processes and remain small, whereas their longitudinal momenta can increase with the quark energy. The nomenclature, then, is a description of a shower of particles confined to a narrow cone about the direction of the primary quarks direction, or a *jet* of particles. Jet production was first observed in e^+e^- colliders in 1975 [93].

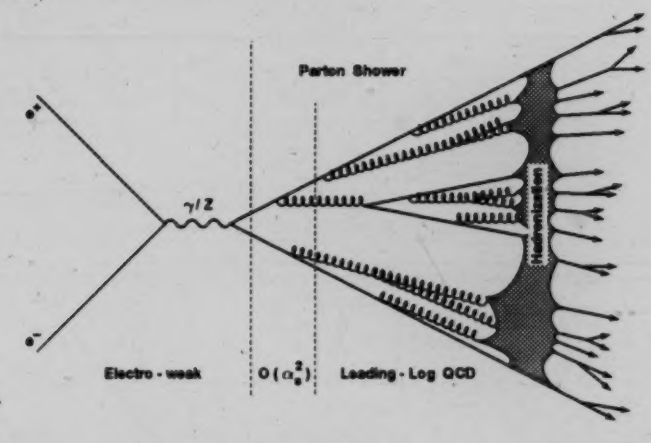


Figure 4.1: Presented is a graphic depiction of the hadronization process.

The process by which the $q\bar{q}$ pairs in the signal and background events evolve into a system of colourless hadrons is known as *hadronization*. Hadronization is dominated by nonperturbative processes which in general cannot be expressed in analytic form.

To make this more quantitative, we need to introduce the concept of a *jet measure*, that is, a procedure for classifying a final state of hadrons according to the number of jets it contains. To be useful, a jet measure should be calculable in perturbation theory and give cross-sections which are free of soft and collinear singularities as well as being relatively insensitive to the non-perturbative fragmentation processes.

One of the first attempts to define jet cross-sections in perturbation theory was by Sterman and Weinberg [94]. In this original attempt, a final state is classified according to the number of non-overlapping cones of half-angle δ required to account for all but a fraction ϵ of the total energy. It turns out that from an experimental and theoretical point of view, the Sterman-Weinberg jet definition is not well suited to analysing multi-jet final states. One of the reasons being the inefficient "tiling" of the 4π solid angle. For this reason various alternatives were proposed, one of the most important being the *minimum invariant mass*, or JADE [95] algorithm, which is characterized by a scale parameter often denoted as the y-cut.

Starting from an n-parton/cluster state, identify the pair with the minimum invariant mass squared. If this is greater than $y_{\rm cut}$, then the number of jets is n, if not then combine the two into a single parton.

$$\min\left[(p_i+p_j)^2\right] = \min\left[2E_iE_j(1-\cos\theta_{ij})\right] > y_{\rm cut}$$

Repeat for the $(n-1)$-parton/cluster state until all parton/clusters have a relative invariant mass squared $> y_{\rm cut}$. According to this definition, an n-parton/cluster state may yield between 2 to n jets. Studies using hadronization models have shown that at high energies, QCD predictions of jet production rates using this jet definition reproduces data well.

Unfortunately, this algorithm tends to reconstruct spurious jets (*ie.*, clusters or hadrons whose momentum does not obviously coincide with any approximately collinear set of partons). To avoid this, the *Durham* [96] algorithm was proposed. This replaces the previous clustering measure by the minimum of the relative transverse momenta,

$$2\min\left[(E_i^2,E_j^2)(1-\cos\theta_{ij})\right] > y_{\rm cut}$$

which also has the property that n-jet production rates can be calculated in perturbation theory.

Because of the striking multi-jet topology of $h^0A^0 \to 3(b\bar{b})$ production, the Durham algorithm was chosen for this analysis. Using the Durham algorithm then, the corrected tracks and clusters were used to reconstruct a six jet topology. Typically, this algorithm is applied by specifying a minimum value for $y^D_{n,n+1}$, where $y^D_{n,n+1}$ is the jet resolution parameter at which events change from n to $n+1$ jets in the Durham scheme. Here, however, events which failed to be reconstructed as six jets with $y^D_{5,6} > 0$ were rejected. The values of $y^D_{n,n+1}$, $n = 2,\ldots,5$ were then used as variables later in the analysis.

The jet 4-momenta were then corrected *via* a 4-C fit [97], enforcing overall energy and momenta conservation. Events where this 4-C fit failed to converge were cut from the analysis.

4.3.6 Topological Variables

To describe and quantify the *jettiness* of an event, a number of variables calculable from the particle distributions, independent of their jet association, are available [83].

For each event we define the *normalized momentum tensor*,

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i \mathbf{p}_i^2} \tag{4.1}$$

where α, β run over the space indices and $\mathbf{p}_i$ is the momentum of the i^{th} particle, summed over all particles in the event. Ideally, all particles would be used but typically only charged particle information is available.

As $S^{\alpha\beta}$ is a real symmetric matrix, it can be diagonalized with eigenvectors $\hat{n}_i$ and eigenvalues λ_i, $i = 1,2,3$, such that $\lambda_1+\lambda_2+\lambda_3 = 1$ and $0 \le \lambda_1 \le \lambda_2 \le \lambda_3$. These eigenvalues can be used to quantify the event shape:

- Roughly Spherical events $\Rightarrow$ $\lambda_1 \approx \lambda_2 \approx \lambda_3$
- Flat (Coplanar) events $\Rightarrow$ $\lambda_1 \ll \lambda_2$
- Cigar (Colinear) events $\Rightarrow$ $\lambda_2 \ll \lambda_3$

Particular combinations of the eigenvalues have been given names. The *Sphericity* [98],

$$S = \frac{3}{2}(\lambda_1+\lambda_2) = \frac{3}{2}\min\left[\frac{\sum_i(\mathbf{p}_i^T)^2}{\sum_i \mathbf{p}_i^2}\right] \tag{4.2}$$

where $\mathbf{p}_i^T$ is the momentum component transverse to the axis which minimizes the numerator. The Sphericity lies in the range $0 \le S \le 1$. Events with $S \approx 1$ are rather spherical, $S \ll 1$ are

back-to-back. In this analysis then, a cut on large S values can be used to reduce the large e^+e^- $(Z/\gamma)^*$ $q\bar{q}$ background.

Other combinations of interest to this analysis are:

- C parameter, $C = 3(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)$, is used to identify 3 jet structures, as it vanishes in 2 jet case.
- D parameter, $D = 27\lambda_1\lambda_2\lambda_3$, is used for 4 jet structures, as it vanishes for planar events.

The C parameter is particularly useful at the preselection level in reducing the $e^+e^- \rightarrow (Z/\gamma)^* \rightarrow f\bar{f}$ background as it rapidly approaches zero for two-jet events. Thus, the C-parameter is required to be ≥ 0.25.

Variables based on $S^{\alpha\beta}$, however, do have some shortcomings. The main drawback is the fact that the particles are weighted by the squares of their momentum in the denominator of $S^{\alpha\beta}$. This has the effect that the occasional energetic track can have an exaggerated effect on the tensor. Furthermore, $S^{\alpha\beta}$ has a number of theoretical problems. From the parton point of view, the quark can split into collinear quark-gluon pairs with a high probability, in fact diverging at the infra-red gluon limit. In this case the final state quark and gluon have the same momentum as the parent quark but contribute quite differently to any quadratic function. $S^{\alpha\beta}$ is therefore intrinsically unstable to such collinear splittings, which are a very common features of QCD radiation. Calculating $S^{\alpha\beta}$ at the hadronic level also runs into the same problem when a decaying hadron is replaced by a number of almost collinear daughter particles.

Variables based on the linear sums of particle momenta are stable against these collinear splittings, so are more insensitive to the fragmentation details in the hadronization of the primary partons. One such variable is the *Thrust* [99], defined by

$$T = \max \frac{\sum_i |\mathbf{n} \cdot \mathbf{p}_i|}{\sum_i |\mathbf{p}_i|} \tag{4.3}$$

where the *thrust axis* is given by the $\mathbf{n}$ which maximizes T. Spherical events correspond to $T \approx 1/2$ and cigar-like events to $T \approx 0$. The thrust axis is a useful vector as it gives an indication of the energy flow in the event. To ensure then that the event is well contained in the sensitive areas of the detector, the polar angle of the thrust axis, θ_T, is required to satisfy $|\cos\theta_T| \leq 0.9$.

One last class of event variables considered in this analysis are due to G. Fox and S. Wolfram [100]. The Fox-Wolfram moments, H_ℓ, $\ell = 0, 1, 2, ...$ are defined as

$$H_\ell = \sum_{i,j} \frac{|\mathbf{p}_i||\mathbf{p}_j|}{E^2_{vis}} P_\ell(\cos\theta_{ij}) \tag{4.4}$$

where θ_{ij} is the opening angle between particles i, j, E_{vis} the total visible energy of the event and P_ℓ are the Legendre polynomials. To the extent particle masses may be neglected, $H_0 = 1$. Otherwise the normalized moments are defined as $H_{\ell_0} = H_\ell / H_0$. If momentum is balanced, $H_1 = 0$. Two-jet events tend to give $H_\ell \approx 1$ for ℓ even and ≈ 0 for ℓ odd. The 2^{nd} moment is linearly related to the C parameter. However, the larger moments provide new topological variables.

4.3.7 Initial State Radiation

One last physical parameter that proved exceedingly valuable in reducing the $e^+e^- \rightarrow (Z/\gamma) \rightarrow f\bar{f}$ background was the identification of energetic Initial State Radiation (ISR). At LEP2 energies, the fast decrease in the cross-section observed in e^+e^- interactions, as the energy moves away from the Z^0 peak, favours the radiation of hard collinear photons which boost the effective centre of mass energy, $\sqrt{s'}$, of the e^+e^- collision back to the Z^0 mass; this is the so called Z^0 *return* [101].

$\sqrt{s'}$ for each hadronic event was estimated as follows [102]. In approximately 20% of radiative events the photon was detected directly in the electromagnetic calorimeter. Such photons were

identified by requiring them to conform to the expected transverse shower shape and to be isolated by observing no more than 1 GeV in a cone of half-angle 200 mrad. In the remaining cases jets were identified using the Durham jet finding scheme, then the measured energy and momentum were used to perform a kinematic fit to a system of jets and an unmeasured photon along the beam direction, imposing constraints of energy and momentum conservation. The photon energy, E_γ, derived from the event kinematics or the energy found in the calorimeter, whichever was the larger, was used to compute $s' = s - 2E_\gamma\sqrt{s}$ [2].

ISR associated with $e^+e^- \to (Z/\gamma)^* \to f\bar{f}$ background tend to generate $\sqrt{s'} \approx m_{Z^0}$ and are efficiently rejected by requiring $\sqrt{s'} \geq 100\,\text{GeV}$.

4.3.8 Summary of Preselection Cuts

As we have taken time to define the standard variables used in multi-hadronic analyses as we came upon them, while also introducing new variables not directly related to the preselection, we will summarize:

- The detector status was acceptable.
- The LEP2 Mutlihadronic (L2MH) event selection flag was set.
- Charged track multiplicity ≥ 25.
- The polar angle of the thrust axis, θ_T, is required to satisfy $|\cos\theta_T| \leq 0.9$.
- The C-parameter ≥ 0.25.
- The effective centre of mass energy, after initial state photon radiation, $\sqrt{s'} \geq 100\,\text{GeV}$.
- 6 jets were found, as discussed on p. 80.

The signal Monte Carlo survived these preselections with better than 90% efficiency and reduced the large $(Z/\gamma)^* \to f\bar{f}$ background by $\approx 90\%$.

Once manageable analysis sets were created, the following selections were applied at analysis level,

- The jets' charged track and electromagnetic cluster multiplicity ≥ 2
- The 4C fit which enforces overall energy-momentum conservation was required to succeed.

These latter two cuts were not subject to optimization as the goal was to ensure good event quality with well defined jets.

4.4 B Jet Tagging

Since neutral Higgs bosons decay preferentially to $b\bar{b}$ pairs, the tagging of jets originating from b-quarks plays an important role in most Standard Model Higgs boson searches. This channel in particular emphasizes the b-quark content in order to discriminate between the signal and $(Z/\gamma)^*$ or W^+W^- multi-hadronic final states.

The existence of hadrons containing heavy quarks is deduced by observations of their decays. Any experiment which depends of identifying the presence of heavy quarks would benefit from at least a phenomenological, if not a theoretical, understanding of heavy quark decays processes. The simplest approach is to treat the decays of hadrons containing heavy quarks in analogy with the

[2] If p_1, p_2 are the initial electron momenta, then by conservation of four-momentum we can write $p_1 + p_2 - p_\gamma = p'_1 + p'_2$, with p'_1, p'_2 the electron momenta after the ISR, p_γ. Squaring both sides, and noting $s = (p_1 + p_2)^2$, we find $s' = s - 2E_\gamma\sqrt{s}$.

decay of a free muon. This is called the *spectator model* [96, 103], since the quarks which accompany the heavy quark in the hadron are assumed to play no role in its decay. After the decay of the heavy quark, *via* virtual W emission, all the remaining quarks form hadrons.

One immediate consequence of the spectator model is that all hadrons containing one heavy quark of a particular flavour should have the same lifetime. For the case of the charm quark, this is not precisely true. For example, for $D^0(c\bar{u})$, $D^+(c\bar{d})$ and $\Lambda_c^+(cud)$ the measured lifetimes are [80]:

$$\begin{aligned} \tau_{D^0} &= 4.76 \pm 0.04 \times 10^{-13}\ \mathrm{s}\ , \\ \tau_{D^+} &= 10.57 \pm 0.15 \times 10^{-13}\ \mathrm{s}\ , \\ \tau_{\Lambda_c^+} &= 2.06 \pm 0.12 \times 10^{-13}\ \mathrm{s}\ , \end{aligned}$$

which are clearly not identical but are at least of the same magnitude. Aside from QCD corrections, however, these complications mainly affect the non-leptonic decay modes; the leptonic decays modes are relatively clean. One should note that the energy release in a c quark decay is only $\mathcal{O}(1\,\mathrm{GeV})$ so we would expect the spectator approximation would improve with the heavier b quark.

At the b quark vertex, $b \to c\mathrm{W}^-$ and $b \to u\mathrm{W}^-$ are kinematically allowed, of which the CKM mixing matrix indicates that $b \to c\mathrm{W}^-$ is dominate. When leptons are present in the decay mode of a hadron, the decay is termed semileptonic; otherwise, the decay is wholly hadronic. Including a colour factor of 3, there are nine principal decay modes: the three semileptonic modes $\mathrm{W}^- \to \ell\bar{\nu}_\ell$, $\ell \in \{e, \nu, \tau\}$ and the six hadronic modes $\mathrm{W}^- \to d\bar{u}, s\bar{c}$. To the accuracy of this discussion, the $\tau\bar{\nu}_\tau$ and $s\bar{c}$ suffer relative phase space suppression of approximately 1/5 and 1/10 respectively.

In the spectator model, the b decay is closely analogous to $\mu \to \nu_\mu e\bar{\nu}_e$ with the charm mass in b decay playing a similar role to the electron mass in μ decay. Hence, we may write,

$$\tau_b \approx \frac{\tau_\mu}{6\,|V_{cb}|^2} \left(\frac{m_\mu}{m_b}\right)^5 \frac{1}{f(\frac{m_c}{m_b})}\ , \tag{4.5}$$

where $f(\epsilon) = (1-\epsilon^4)(1-8\epsilon^2+\epsilon^4) - 24\epsilon^4 \ln\epsilon$ is the phase space factor. *A priori*, one may expect the B hadron lifetime to be a factor of $(m_c/m_b)^5$ shorter than the estimate for charm flavoured hadrons, as follows from the previous equation. This mass effect, however, is almost entirely canceled by the CKM factor $|V_{cb}|^2$.

If we take [103] $m_b = \frac{1}{2}m_\Upsilon \simeq 4.73\,\mathrm{GeV}$ and $m_c = \frac{1}{2}m_{\psi} \simeq 1.55\,\mathrm{GeV}$, we obtain an estimate for τ_b of $1.8 \times 10^{-12} \times (0.05/|V_{cb}|)^2$ s. Estimating $V_{cb} \approx 0.05$, from the observed $B^\pm$ and B^0 semileptonic partial widths, we find our final estimate, $\tau_b \approx 1.8 \times 10^{-12}$ s which compares favourably to the observed average B hadron lifetime [80], $\tau_{B_d^0} = 1.56 \pm 0.04 \times 10^{-12}$ s, supporting the spectator model assumptions in b quark systems. Likewise, the ground-state hadrons containing b quarks have roughly equal lifetimes, further supporting the spectator model. The observed B hadron lifetime corresponds to a proper lifetime expressed in units of length of $c\tau = 468\,\mu\mathrm{m}$. A b quark with a momentum of 20 GeV, a typical jet energy at OPAL will have a relativistic γ factor of about 4, in which case the corresponding B hadron will travel $\mathcal{O}(2\,\mathrm{mm})$. Such a decay length will produce a secondary vertex, as depicted in Figure 4.2, which in OPAL is reconstructed using information from the central tracking chambers in conjunction with the silicon microvertex detector.

Returning to the muon decay analogy, it can be shown that the decay electrons have a hard energy spectrum. That is, in the muon rest frame the electron energy is a significant fraction of the muon rest mass. A similar feature exists for the semileptonic decays of the b quark, where the leptonic energy spectrum of $b \to e$ and $b \to \mu$ are essentially the same since $\frac{m_e}{m_b} \simeq \frac{m_\mu}{m_b} \ll 1$.

While the c flavoured hadrons have been shown to vary more from the spectator model predictions than the B hadrons, the basic phenomenology is still the same. So the decay of a charmed hadron through virtual W emission can generate a displaced secondary vertex, larger jet multiplicity and in the semileptonic case a hard lepton. When the charmed hadron appears in a cascading decay chain of a primary B hadron, through $b \to c \to d, s$, this will contribute to the identification

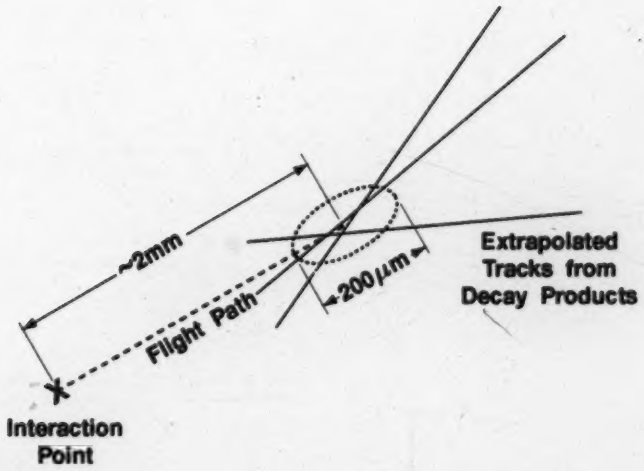


Figure 4.2: Typical dimensions of a B-hadron decay at OPAL. Lifetime tagging through secondary vertex reconstruction.

of b flavoured jets, otherwise, primary c production will be a source of irreducible background to any b tagging methods.

In summary, the spectator model of B hadron decays predicts,

- All hadrons containing a b quark will decay with similar lifetimes.
- At typical OPAL jet energies, these lifetimes correspond to a decay length $\mathcal{O}(2\,\mathrm{mm})$, large enough to be measured in an experiment equipped with a microvertex detector.
- b flavoured jets will have larger charged track multiplicities due the hadronic contributions from the virtual W bosons in the cascading decay chain involving $b \rightarrow c \rightarrow \mathrm{d,s}$.
- The decay of the virtual W contributes to the jet charged track multiplicities through its hadronic decay modes, or, in the leptonic case, produce easily identifiable hard leptons.

In order then to study b quark production at OPAL an algorithm will be presented in the next section which exploits these properties.

4.4.1 Jet-wise B-Tagging Algorithm

In order to have the highest possible efficiency, a jet-wise b-tagging algorithm common to the OPAL Higgs analyses has been developed [104]. While this method was not developed by the author, its results are of such critical importance to the analyses presented in this thesis that the algorithm will be presented in some detail.

The method employs three independent b-tagging methods: (1) lifetime tag, (2) high-p_t lepton tag, and (3) jet shape tag. These three methods, described below, are combined using a Bayesian likelihood method to form a discriminator for each jet. Figure 4.3 schematically illustrates the b-tagging procedure.

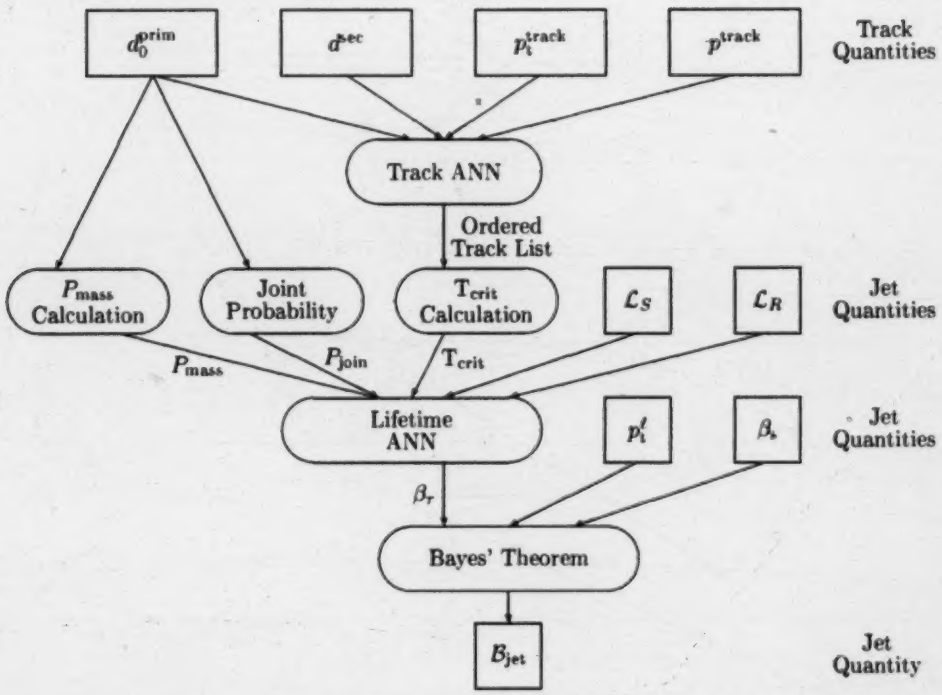


Figure 4.3: *Schematic presentation of the calculation of the jet-wise b-tagging variable $\mathcal{B}_{jet}$. The symbols and algorithms are explained in the text.*

Lifetime tag

The lifetime tag exploits the relatively long lifetime, high decay multiplicity and high mass of the b-flavoured hadrons. Five quantities, based on these properties, are calculated from the tracks and clusters assigned to a given jet. These five quantities are input to an artificial neural network (ANN) to form a lifetime tag, β_τ, for each jet considered.

Three of these quantities rely on the reconstruction of secondary vertices. Within a jet, sub-jets are formed using a cone algorithm [105] with a cone half angle of 0.5 radians and a minimum sub-jet energy of 7 GeV. In each of these sub-jets, secondary vertices are reconstructed using the method described in [106]. An attempt is made to fit all tracks in a sub-jet to a common vertex. Tracks are discarded by an iterative procedure which drops the track with the largest contribution to the χ^2 of the vertex fit, until the largest χ^2 contribution is less than 4, with at least two tracks remaining.

The following quantities are calculated as inputs to the ANN:

- Secondary vertex likelihood, $\mathcal{L}_S$: $\mathcal{L}_S$ is calculated from the probability density function of the decay length significance, S (the decay length divided by its error) for b, c and uds flavours. If more than one sub-jet is formed, the secondary vertex with the largest $\mathcal{L}_S$ in a given jet is selected.

- Reduced secondary vertex likelihood, $\mathcal{L}_R$: the reduced decay length is obtained from a vertex fit using all tracks in the secondary vertex, except the one with the largest impact parameter significance (i.e., the impact parameter with respect to the primary vertex divided by its error). For b-flavoured hadron decays, the reduced decay length often coincides with the

decay length, however, randomly formed vertices exhibit less robustness against removing the most significant track. The reduced decay length significance R is given by the reduced decay length divided by its error.

- Critical track discriminator, T_{crit}: An auxiliary ANN is trained to discriminate between tracks originating from the b-flavoured hadron decay and from tracks due to fragmentation or decays of light-flavoured hadrons. The inputs to this ANN are the impact parameter of the track with respect to the primary vertex, d_0^{prim}; the impact parameter with respect to the secondary vertex, d^{sec}; the momentum of the track, p^{track}; and its transverse momentum with respect to the corresponding sub-jet axis, p_t^{track}. The tracks belonging to the sub-jet are then sorted in descending order according to the output of the auxiliary ANN. Tracks are added one by one to a 'cluster' of tracks whose invariant mass is calculated, assuming that all tracks have the pion mass. T_{crit} is the auxiliary ANN output of the track which causes the cluster invariant mass to exceed 1.9 GeV. This algorithm [107] exploits the higher mass of b-flavoured hadrons compared to charmed and lighter hadrons.

Not all b-flavoured jets produce identifiable secondary vertices. Thus, in order to compensate for the loss in b-tagging efficiency due to the requirement of secondary vertex reconstruction, two of the inputs to the ANN are based on track impact parameters only.

- Two-dimensional impact parameter joint probability, (P_{join}): The impact parameter distribution for tracks with negative impact parameter significance is assumed to come from the primary vertex and, thus, provides an estimate of the detector resolution function [3]. The resolution function is then used to weight the tracks, and the joint probability for the tracks in a jet to come from the primary vertex is given by

$$P_{join} = y\Sigma_{m=0}^{N-1}\frac{(-\ln y)^m}{m!},$$

where y is the product of the probabilities of all N tracks in the jet [108].

The five variables $\mathcal{L}_S$, $\mathcal{L}_R$, T_{crit}, P_{join} and P_{mass} are then input to an artificial neural network. Monte Carlo samples at $\sqrt{s} = m_{Z^0}$ were used to train the ANN, with one output node, the lifetime-tag β_τ. The vertex tagging performance was found to depend on the jet polar angle, so three separate ANNs are trained for jets with $|\cos\theta_{jet}| \leq 0.75$, $0.75 < |\cos\theta_{jet}| \leq 0.9$, and $|\cos\theta_{jet}| > 0.9$.

The high-p_t lepton and jet shape tags

Semileptonic b-decays are identified using electron and muon selections, rejecting electrons from γ conversions as described in [109]. The transverse momentum, p_t^ℓ, of the lepton, calculated with respect to the direction of the sub-jet which includes the lepton track, is used as a b-tag variable.

The larger decay multiplicity and higher mass of the b-flavoured hadrons tend to result in a more spherical shape for b-jets compared to lighter flavour jets. As a measure of the jet shape, the boosted sphericity β_s, defined as the sphericity (page 80) of the jet calculated in its rest frame, is used as the jet shape tag.

4.4.2 Application to the Six Jet Analysis

For a given jet then, we calculate the three quantities β_τ, p_t^ℓ, and β_s as described above. Comparisons of Monte Carlo and Data can then be used to define the probability density functions, $P(t|q)$, as the probability of observing a value t, for the tagging method $t \in \{\beta_\tau, p_t^\ell, \beta_s\}$ for each jet flavour $q \in \{b, c, uds\}$.

[3] The impact parameter is taken to be positive if in the two-dimensional projection the track path, starting from the point of closest approach to the primary vertex, crosses the jet axis in the flight direction; otherwise it is negative.

Of interest now is the subsequent classification of the jet's flavour based on these observations, that is, given $(\beta_\tau, p_t^\ell, \beta_s)$ what is the probability the jet is b-flavoured? If we denote $P(\beta_\tau, p_t^\ell, \beta_s|b)$ as the conditional probability of observing $(\beta_\tau, p_t^\ell, \beta_s)$ for a b-flavoured jet, then Bayes theorem [110] tells us

$$P(b|\beta_\tau, p_t^\ell, \beta_s) = \frac{P(\beta_\tau, p_t^\ell, \beta_s|b)P(b)}{P(\beta_\tau, p_t^\ell, \beta_s)} . \tag{4.6}$$

Since the choices of jet flavour classification are exhaustive, $\bigcup_{q=b,c,uds} q = \Omega$, and mutually exclusive, $q_i \bigcap q_j = \emptyset, \forall i,j \in \{b,c,uds\}$, the probability of the event $(\beta_\tau, p_t^\ell, \beta_s)$ can be written as

$$P(\beta_\tau, p_t^\ell, \beta_s) = \sum_{q=b,c,uds} P(\beta_\tau, p_t^\ell, \beta_s|q)P(q) \tag{4.7}$$

allowing us to write Equation 4.6

$$P(b|\beta_\tau, p_t^\ell, \beta_s) = \frac{P(\beta_\tau, p_t^\ell, \beta_s|b)P(b)}{\sum_{q=b,c,uds} P(\beta_\tau, p_t^\ell, \beta_s|q)P(q)} . \tag{4.8}$$

The factors $P(q)$ in Equation 4.8 are interpreted as the absolute probability of a jet being q-flavoured which is, in turn, related to the cross-sections and branching fractions of the processes contributing to the hadronic jets observed in data. These absolute probabilities are to be measured from the flavour compositions in the contributing background channels and are herein parametrized by the weight factors w_b, w_c, and w_{uds}.

Further, since the three quantities described above, β_τ, p_t^ℓ, and β_s exploit different properties of b-flavoured hadron decays and are almost uncorrelated, we can factor $P(\beta_\tau, p_t^\ell, \beta_s|q)$ as the product

$$P(\beta_\tau, p_t^\ell, \beta_s|q) = P(\beta_\tau|q) \cdot P(p_t^\ell|q) \cdot P(\beta_s|q) \tag{4.9}$$

where each term in the expansion is known from the previous discussion.

The final b-tagging discriminant for each jet is defined as

$$\begin{aligned} \mathcal{B}_{\text{jet}} = P(b|\beta_\tau, p_t^\ell, \beta_s) &= \frac{w_b \cdot P(\beta_\tau|b) \cdot P(p_t^\ell|b) \cdot P(\beta_s|b)}{\sum_{q=b,c,uds} w_q \cdot P(\beta_\tau|q) \cdot P(p_t^\ell|q) \cdot P(\beta_s|q)} \\ &= \frac{1}{1 + \sum_{q=c,uds} \frac{w_q}{w_b} \cdot \frac{P(\beta_\tau|q) \cdot P(p_t^\ell|q) \cdot P(\beta_s|q)}{P(\beta_\tau|b) \cdot P(p_t^\ell|b) \cdot P(\beta_s|b)}} \\ &= \frac{1}{1 + \alpha\mathcal{C} + \beta\mathcal{U}} \end{aligned} \tag{4.10}$$

with $\alpha = w_c/w_b$, $\beta = w_{uds}/w_b$ and $\mathcal{C}$, $\mathcal{U}$ are the respective ratios of the jets probability of being c or uds-flavoured with respect to its probability of being b-flavoured. In this format, we see there are only two free parameters, α, β, related to the unknown flavour weights, which could be fit to each analysis based on efficiency and purity requirements.

The obvious way to define a b-tagging discriminant for N-jet event is $\mathcal{B}_{\text{Event}} = \prod_{i=1}^{N} \mathcal{B}_{\text{jet}_i}$, where the product is indexed by the jet number. This would work well when the ratios α and β are known, but if not then the resulting product could be very sensitive to their selected values and will be too non-linear to fit to data. For this analysis, we treat the factors α, β as unknown, and define the associated, N-jet, event b-tagging discriminant is

$$\mathcal{B}_{\text{Event}} = \frac{1}{1 + \alpha \prod_{i=1}^{N} \mathcal{C}_i + \beta \prod_{i=1}^{N} \mathcal{U}_i} . \tag{4.11}$$

If we present only the $\mathcal{C}_i$, $\mathcal{U}_i$ products as inputs to a neural network, we would then expect the network to select the optimal values for α, β to classify the B'ness of signal and background events.

Thus, we define the two input variables

$$\texttt{CLIKE} = \prod_{i=1,4} \mathcal{C}_i \,, \quad \texttt{ULIKE} = \prod_{i=1,4} \mathcal{U}_i \,, \tag{4.12}$$

as the product of the $\mathcal{C}$, $\mathcal{U}$ values for the 4 highest energy jets. This is done to reduce the effect of poorly reconstructed low energy jets in the background.

4.4.3 Secondary Vertices

Before the development of the jet-based b-tagging method described in the previous sections, researchers relied heavily on the direct reconstruction of secondary vertices to indicate the presence of long lived b hadrons. At the time this analysis was performed, the jet-based b-tagging method was not well tested, so the author felt it prudent to contrast it with the discriminating power of standard vertex based variables.

Secondary vertices were constructed in each jet using the *Build Up* [111] (denoted BU) and *Tear Down* (denoted TD) [111, 106] methods.

The BU vertex finding method starts by trying to form a nucleus secondary vertex from tracks with significant impact parameters with respect to the primary vertex. Tracks are added to this nucleus if they are closer to it than the primary, and again the separation from the primary can be used to tag b jets. The BU vertices were found to contribute little information to neural networks which contained TD vertex information and so are not used in this analysis.

The TD method starts by forming a common vertex using all the tracks in the jet, and progressively removes tracks which do not fit well. When all the remaining tracks form a good vertex, the separation from the primary vertex is used to distinguish b from non-b jets. Accurate secondary vertex reconstruction is achieved by requiring that a minimum number of tracks assigned to the vertex have well defined hits in the Silicon Microvertex Detector. A good(reduced) tear down secondary vertex is defined as having $\geq 2(3)$ tracks each with ≥ 2 $r\phi$ and ≥ 1 rz silicon hits, using the tear down vertex reconstruction algorithm. As mentioned previously, the *decay length significance* is defined as the decay length divided by its error.

For robustness, the *reduced* decay length significance is defined as the decay length significance calculated for the vertex after the track with the highest impact significance had been removed. For b-flavoured hadron decays, the reduced decay length often coincides with the decay length, randomly formed vertices exhibit less robustness against removing the most significant track. In this case we originally require 3 good tracks so that the vertex will still be well defined.

Following our expectation that b-flavoured jets will optimally return secondary vertices with positive significance, we define the significance and reduced significance sum for an event by adding the significances of all the good(reduced) secondary vertices with positive significance which we will denote by TSUM(TRSUM). Although highly correlated, a plot of Significance Sum vs. Reduced Significance Sum was not perfectly linear and so includes discriminatory information. This was directly confirmed by comparing ANN trained with and without one or the other of these inputs.

One may notice that information from the secondary vertices in the jets is used in determining the $\mathcal{C}$, $\mathcal{U}$ values described above. However, Figure 4.4 shows a comparison between TSUM and the CLIKE, ULIKE variables where we see that the relationship is not linear and, thus, these additional variables contain new discriminatory information.

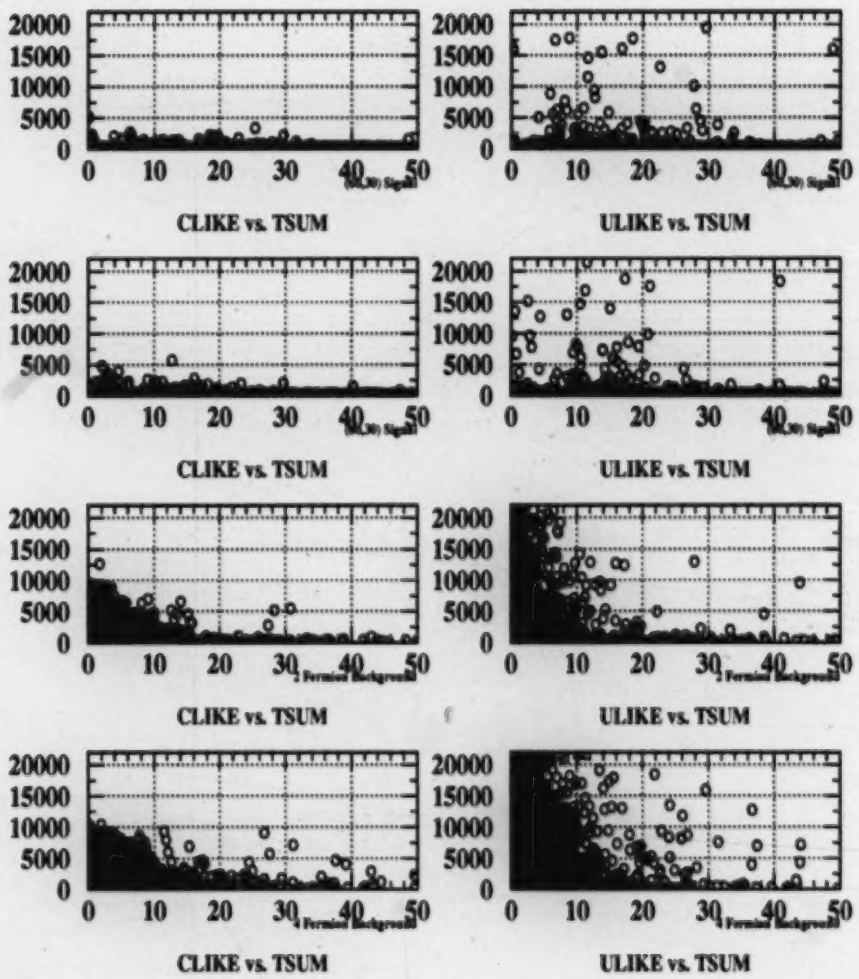


Figure 4.4: Comparison between between TSUM and the CLIKE, ULIKE variables. The data types are indicated along the horizontal axis of each plot. We see that the relationship is not linear and that there exists discriminatory information between them.

4.5 Artificial Neural Network Analysis

The $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV signal was found to present quite a different topology compared to that of the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV signal and more closely resembled the $q\bar{q}(g)$ background. Thus, the network was trained separately on the signal events $hA \to 3(b\bar{b})$, with $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV and $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV. We will refer to these two networks as $a63$ and $a83$, respectively. The $(m_{h^0}, m_{A^0}) = (80, 40)$ GeV case was not included in the training as it seemed to have a very similar topology to the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV case. This seems to be supported by similar final efficiencies for the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV and $(80, 40)$ GeV cases.

4.5.1 Topology and Training

We use a simple feed-forward neural network with backpropagation updating programmed using the `JETNET-3.5` [74] package, with a topology defined by 15 input nodes, 10 hidden nodes and 1 output node.

It is generally accepted that a single hidden layer is sufficient for classification oriented neural networks [74]. Using more hidden layers was not investigated for this analysis. The output node was trained to yield an *a priori* bayesian probability of 0 or 1 depending on whether the event was classified as signal or background. A sigmoidal activation function was employed and the networks were trained with the following common parameter values; learning parameter $\eta = 0.001$, momentum $\alpha = 0.5$, temperature $T = 1.0$. The weights were updated every 10 training patterns processed. Variation of these parameters within reasonable bounds had no significant effect on the efficiency of the net.

The ANN was trained using those events which passed the preselections described in Section 4.3.8. In this way the background rejection is maximized by ensuring that artifacts in the raw Monte Carlo, filtered by the preselection, will not effect the training of the network.

4.5.2 Training and Testing data sets

The available Monte Carlo was split into two subsets: The training data, which was used to train and validate the networks, and the analysis data, used determine the efficiencies quoted as the results of this analysis.

The training sample, listed in Table 4.5, was composed of the large $5K$ signal samples and a subset of the `PYTHIA, GRC4F` [4] background. This training subset was further subdivided into 90% for training phase and 10% for the verification phase.

To help reduce statistical errors of the final signal efficiencies, the 10% of the signal training set used for verification was re-used in the main analysis, effectively halving the statistical error of the relevant signal samples.

During training the background data were mixed based on their relative effective cross-sections (cross-sections multiplied by preselection efficiency). In this case, only the channel with the highest effective cross-section would use 90% of its available data for the training phase. The remaining channels contribute based on their effective cross-section relative to the highest one, with all the remaining data in these channels being used in the verification set. This mixing ensures that all the available Monte Carlo data can be used for training purposes, while ensuring that the networks are not biased by channels which have large sample sizes but physically low relative cross-sections. This prevents the ANN from learning, for example, the $e^+e^- \to q\bar{q}e^+e^-$ background channel at the expense of efficiency in tagging $e^+e^- \to q\bar{q}q\bar{q}$ or $e^+e^- \to q\bar{q}\ell^+\ell^-$ when in actuality the $e^+e^- \to q\bar{q}e^+e^-$ contributes very little to the Standard Model predicted background.

[4] $e^+e^- \to q\bar{q}e^+e^-$ data was not used in training as its effective cross-section was too low to contribute to the final expected background.

Process	Run	Events	σ (pb)
HZHA generated signal, $h^0A^0 \to 3(b\bar{b})$			
mh=60, mA=30	R7048	5K(3961)	
mh=80, mA=30	R7049	5K(4260)	
Main PYTHIA, GRC4F background			
$Z^0(\gamma) \to q\bar{q}$	R5050	93.5K(5426)	107.43(6.23)
$e^+e^- \to q\bar{q}q\bar{q}$	R7051	10K(5887)	7.86(4.63)
$e^+e^- \to q\bar{q}\ell^+\ell^-$	R7050	40.6K(1307)	8.11(0.26)

Table 4.5: Monte Carlo signal and PYTHIA, GRC4F background data sample used in the training of the ANNs. The values in the Events column indicate the total sample size considered before preselection, those in parentheses indicate the number of events passing the preselections and thus are the total number used in the training of the ANN. The cross-sections in parentheses are the effective cross-sections used to mix the background sets during training.

4.5.3 Network Inputs

The ANNs in this analysis are based on the following 15 inputs, which can be loosely classed into three categories:

B-Tagging To exploit the presence of the many b flavoured jets expected in a signal event, we use a number of B-Tagging methods. The first method is based on the probabilistically defined event b-discriminate, Equation 4.11, which lead to the definition of the CLIKE, ULIKE variables, Equation 4.12. The second method was based on the direct analysis of the secondary vertices reconstructed using the teardown algorithm. It was noted that both methods are in essence highly correlated and so perhaps one could be pruned from the ANN. However, comparing ANNs trained with and without both CLIKE, ULIKE variables and the significance sum variables, TSUM, TRSUM showed that both methods contribute information and one cannot be removed in favour of the other without a loss in discriminating power.

Summarizing then, the B-tagging variables used are:

- CLIKE, ULIKE, the product of the $\mathcal{C}$, $\mathcal{U}$ values for the 4 highest energy jets.
- Tear down significance sum, TSUM
- Reduced tear down significance sum, TRSUM

Topology Relying on b-tagging alone was not sufficient due to a number of processes which can contribute an irreducible background. These are multi-jet events with high b-hadron content and include processes such as $(Z/\gamma)^* \to b\bar{b}$ with hard gluon radiation and $W^+W^- \to q\bar{q}\ q\bar{q}$ where heavy c-flavoured jets are falsely tagged as b-flavoured. The topology, however, of true two or three jet events, such as $(Z/\gamma)^* \to q\bar{q}(g)$, can be quite different compared the six-jet signal.

Standard variables such as the sphericity tensor derived C and D parameters along with the Fox-Wolfram moments were found to have distributions which contain information useful for discriminating between the signal and background channels. Further, the jet reconstruction *y-cut* parameters needed to classify the event as 2, 3, 4, 5 and 6 jets have also been shown to contain useful information.

The following kinematic variables were used:

- The effective centre of mass energy, $\sqrt{s'}$.
- The visible energy of the event, E_{vis}.

- The event Thrust, denoted THRU.
- C and D parameters.
- Logarithms of y_{45}^D, y_{56}^D. $y_{23,34}^D$ were pruned with no serious effect on the ANN.

Hemispheres Related to event topology, considering opposite thrust hemispheres provides some new information. In the cases $m_{h^0} + m_{A^0} << \sqrt{s'}$ we imagine the initial hA pair is created back to back with the A in one thrust hemisphere and the $h \rightarrow AA$ in the other.

In this case, we would expect the charged track multiplicity between the two hemispheres, as well as the reconstructed the invariant mass of each hemisphere, to have an approximately 2 : 1 ratio.

The following input variables were used,

- Smallest of the hemisphere invariant masses, denoted AMASS, identified with the $A \rightarrow b\bar{b}$ decay.
- Largest of the hemisphere invariant masses, denoted HMASS, identified with the $h \rightarrow AA$ decay.

Figures 4.5 - 4.6 present the comparisons between OPAL data and the Monte Carlo background for these variables.

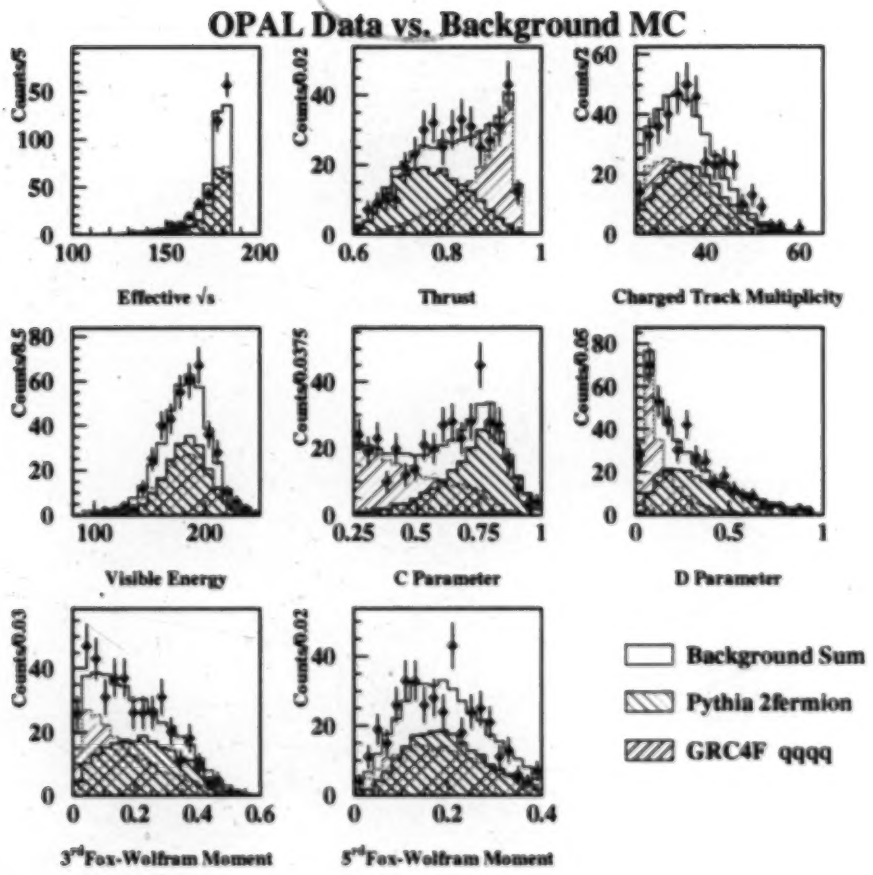


Figure 4.5: Distribution of the Neural Network input parameters for OPAL data (full circles) and the SM prediction (open histogram). The main contributions to the SM background are overlaid as hatched histograms.

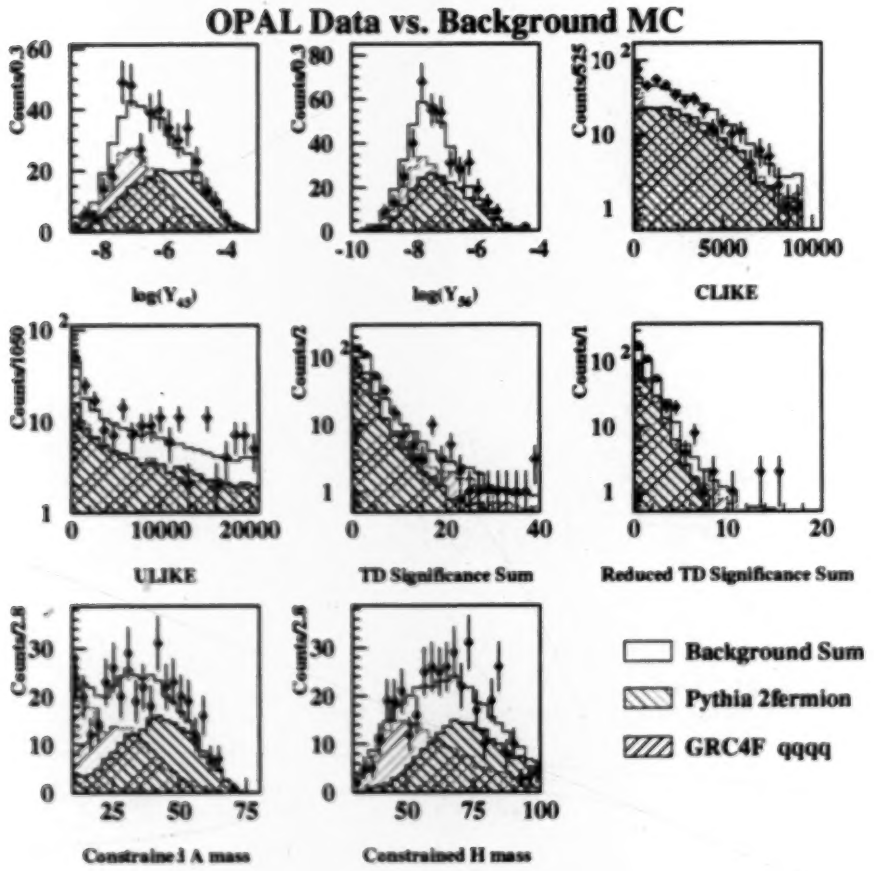


Figure 4.6: Distribution of the Neural Network input parameters for OPAL data (full circles) and the SM prediction (open histogram). The main contributions to the SM background are overlaid as hatched histograms.

Regulating Inputs

To avoid the saturation of the weights in the ANN due to inputs whose magnitude may be orders of magnitude larger than others, the i inputs, x_i are shifted by their means, $\bar{x}_i$, and then scaled by their standard deviations σ_i, ie. $x_i \mapsto \frac{x_i - \bar{x}_i}{\sigma_i}$. The means and standard deviations were calculated using the Standard Model backgrounds only, correctly mixed by their relative cross-sections, which has the benefit that the resulting means could be directly compared to data. Further, to avoid the possibility that the tails in some of these distributions could still be large, we transform as $x_i \mapsto \tanh\left(\frac{x_i - \bar{x}_i}{\sigma_i}\right) \in [-1, 1]$.

Figures 4.7 - 4.8 show the transformed distributions which were used in the ANN training.

Pruning

The original pool of possible inputs to the ANN was selected by highlighting certain aspects of the signal, such as topology, b-tagging and hemisphere related variables. Pruning of the input variables was performed by the time consuming, yet infallible, method of directly comparing networks trained with differing sets of inputs. Figure 4.9 directly compares some of the networks considered by plotting $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV efficiency of these networks for a given expected background level.

At the time of this analysis, the Principal Component Pruning described in a previous chapter had not yet developed, so the number of units in the hidden layer were pruned by directly comparing networks trained with differing number of hidden units.

Network Selection

Once the inputs and network topology had been determined, a choice had to be made as to what stage in the training of the networks was to be selected for the final analysis.

To avoid over-training, the method of early stopping was employed. The networks were trained for a large number of epochs and the generalization error recorded. Networks were considered trained when this error reached its minimum. The selected networks for this analysis attained their minimum generalized error after ≈ 2000 epochs. The evidence for over-training was observed after 5000 epochs.

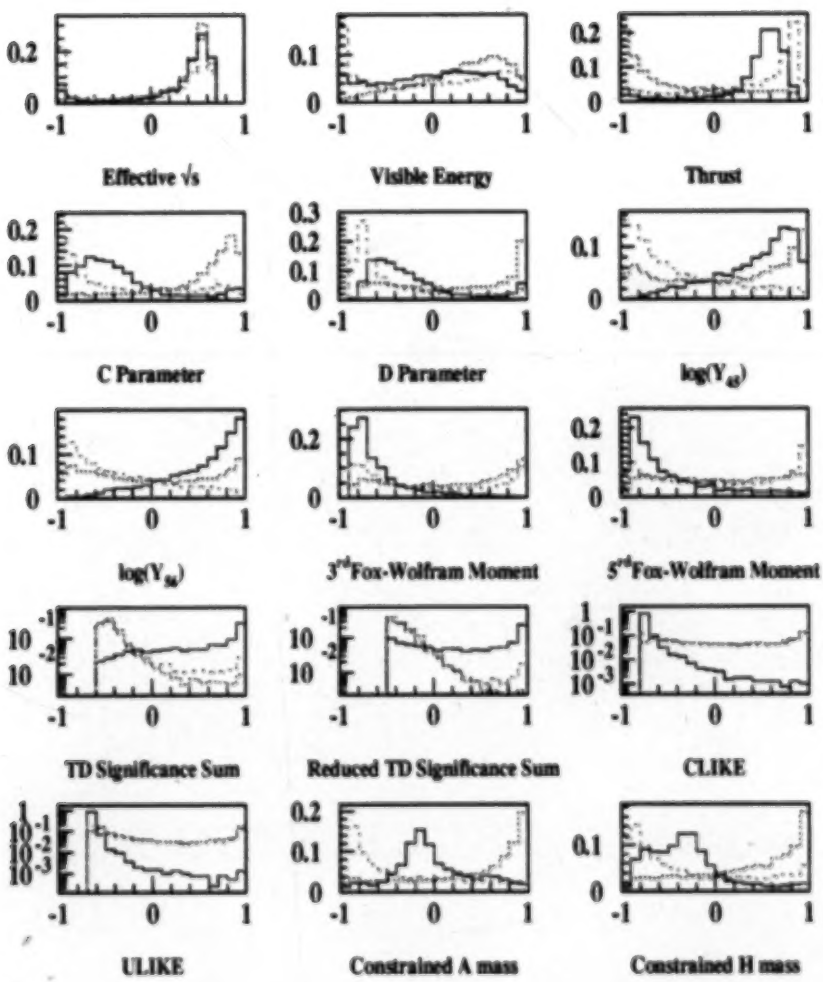


Figure 4.7: Presented are the signal (solid line), Pythia $q\bar{q}$ (dashed line) and GRC4F four fermion (dotted line) input distributions used in the training of the a63 ANN, after application of the $x_i \mapsto \tanh\left(\frac{x_i - \bar{x}_i}{\sigma_i}\right)$ transformation. Distributions are normalized to unity for ease of comparison.

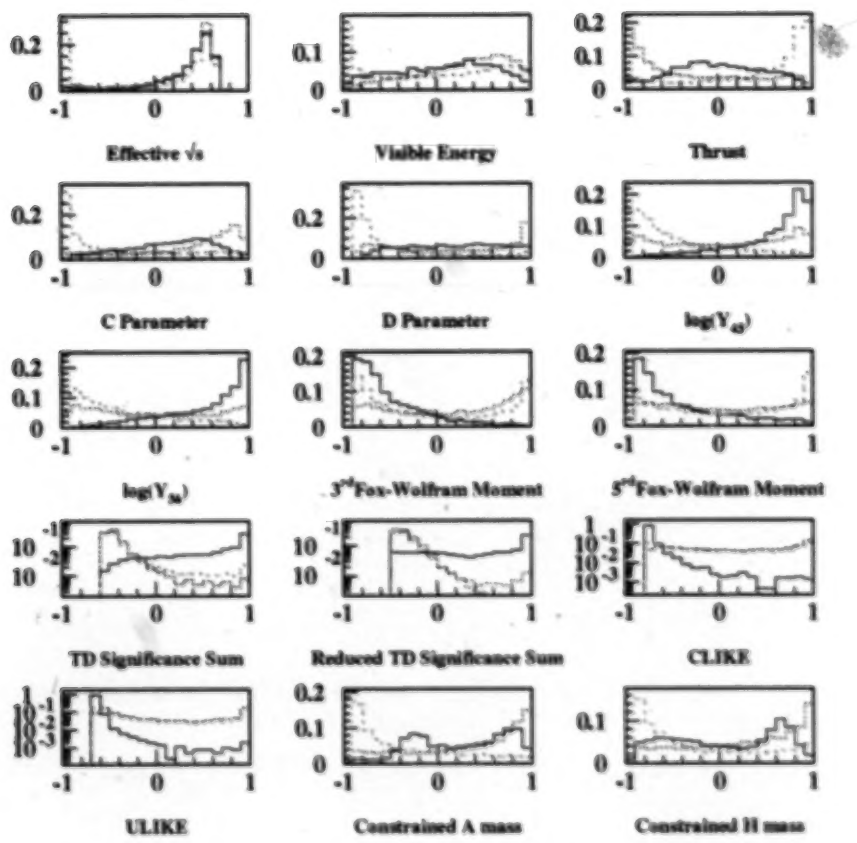


Figure 4.8: Presented are the signal (solid line), Pythia $q\bar{q}$ (dashed line) and GRC4F four fermion (dotted line) input distributions used in the training of the a83 ANN, after application of the $x_i \mapsto \tanh\left(\frac{x_i - \bar{x}_i}{\sigma_i}\right)$ transformation. Distributions are normalized to unity for ease of comparison.

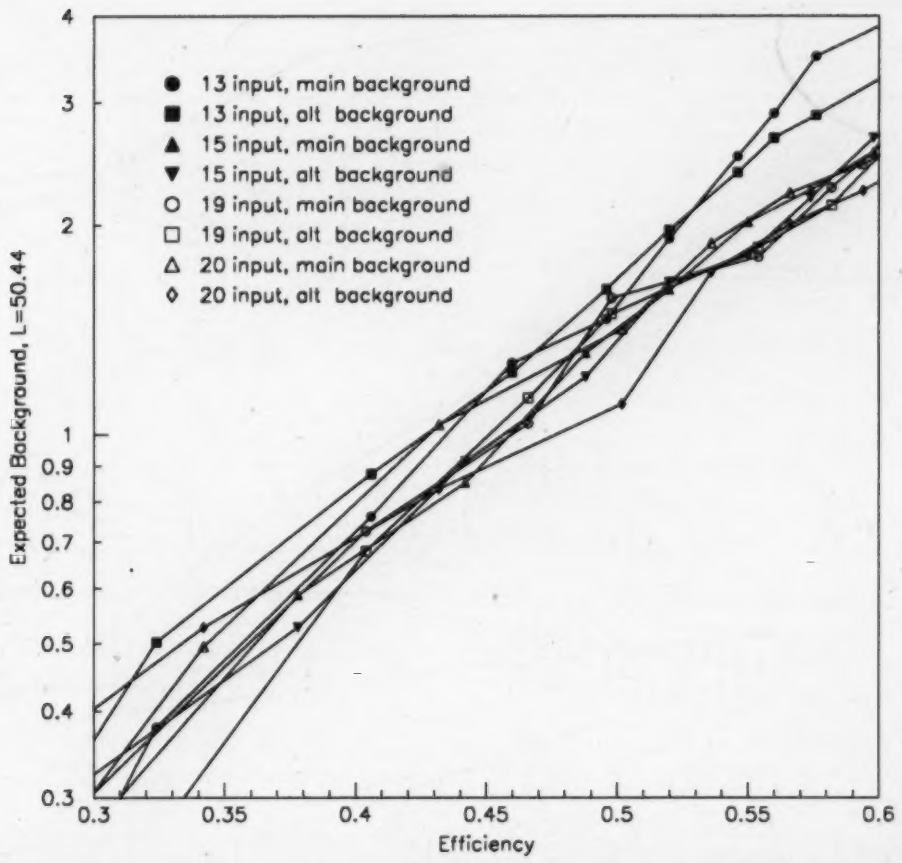


Figure 4.9: ANN ϵ_{signal} vs. Expected Background for a83 trained network. One can see (possibly) that the 20 input ANN has a larger variation between main and alternative background sets. The 15 input ANN is quite similar to the 19 input and the 13 input shows more significant variation.

4.6 Systematic Checks

4.6.1 Preselection

The cuts on the pre-selection parameters $\sqrt{s'}$, charged track multiplicity and C-parameter were varied by an amount equal to the difference between their distributions in data and monte carlo, but had no effect on the primary results.

4.6.2 Si Resolution Smearing and Hit Dropping

To estimate the $r\phi$ and z track resolution uncertainties due to imperfection in the Monte Carlo modelling, the parameterized track smearing routine `ODSMGL` with smearing scale parameters between 0.9 and 1.1 were applied, simulating a 10% reduction and improvement in track resolution. Some relevant distributions are shown in Figure 4.10. The observed results are tabulated in Tables 4.8 and 4.9 and are summarized in Table 4.6.

ANN	Smearing of track resolution				
Analysis	Smear Factor	$\epsilon_{(80,30)}(\Delta)$	$\epsilon_{(80,30)}(\Delta)$	$\epsilon_{(80,40)}(\Delta)$	$\langle n \rangle(\Delta)$
a63	1.1	0.430(0.013)			1.24(0.23))
	0.9	0.430(0.013)			0.91(0.10)
a83	1.1		0.468(0.009)	0.442(0.002)	1.26(0.25))
	0.9		0.456(0.003)	0.424(0.020)	1.07(0.06)

Table 4.6: Results for various smearing values are given for the two neural networks in this analysis. The bracketed values then give the difference between a particular value compared to the primary result which used a smearing of 1.0 .

As well, the parametrized Silicon Microvertex detector (Si) hit dropping in $r\phi$ was varied to measure the analysis' sensitivity to the Si hit matching efficiency. Data showed that the hit matching efficiency may have been slightly overestimated by the 0.08 value used to generate the primary background samples therefore this value was increased by 2% to 1.0 to estimate the error [87]. The complete results are presented in Tables 4.8 4.9 and are summarized in Table 4.7.

ANN	Si hit dropping				
Analysis	Hit Dropping	$\epsilon_{(80,30)}(\Delta)$	$\epsilon_{(80,30)}(\Delta)$	$\epsilon_{(80,40)}(\Delta)$	$\langle n \rangle(\Delta)$
a63	1.0	0.434(0.017)			0.99(0.02))
a83	1.0		0.428(0.031)	0.436(0.006)	0.90(0.11)

Table 4.7: Results for and increase in Si hit dropping in $r\phi$ from 0.8 to 1.0 for the two neural networks in this analysis. The bracketed values then give the difference between a particular value compared to the primary result.

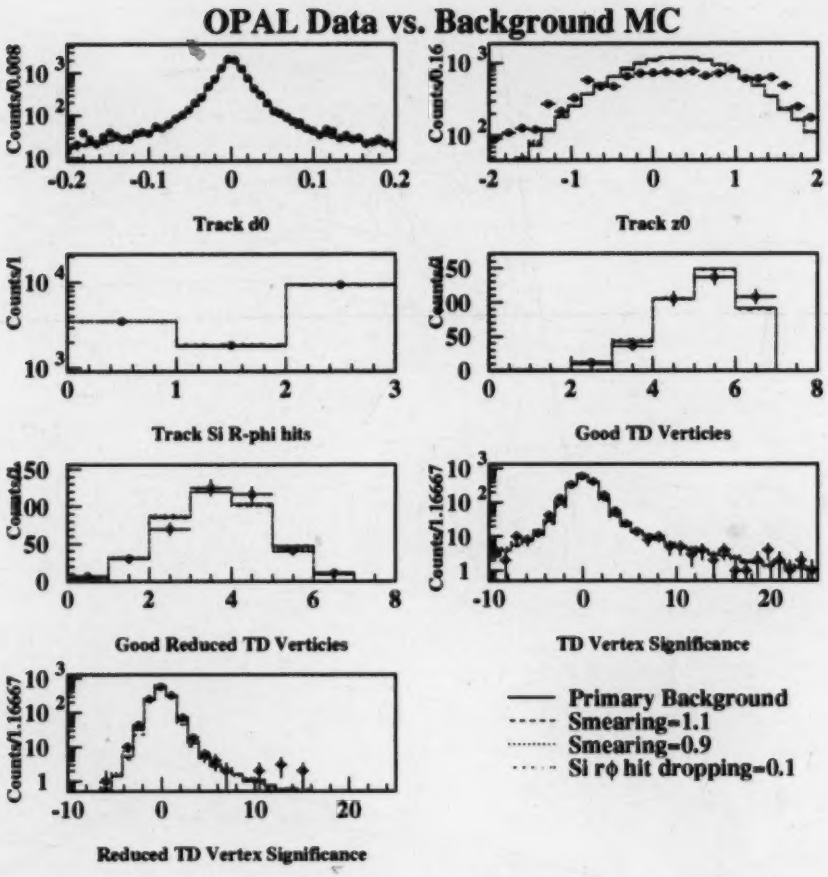


Figure 4.10: Here we see the effect of the various track resolution systematic checks discussed in Section 4.6.2. Data (crossed) is compared with the primary background (solid). The systematic checks are overlaid as indicated in the key.

	data	$hA \to AAA$ (60,30)	(70,20)	(80,30)	(80,40)	$hZ \to ZAA$ (60,30)	(70,20)	(80,30)	(80,40)	$e^+e^- \to$ $q\bar{q}(g)$	$q\bar{q}q\bar{q}$	$q\bar{q}\ell^+\ell^-$	$q\bar{q}e^+e^-$	$\langle n \rangle$
1. Variation in Si hit matching efficiency														
ANN cut	1	217	111	78	32	13	5	10	10	25	9	0	0	
ϵ		.434	.222	.156	.064	.026	.010	.020	.020	.000126	.000600	.0	.0	
$\pm$		.022	.019	.016	.012	.007	.004	.006	.006	.000025	.000200	.0	.0	
$\langle n \rangle$										0.74	0.26	.0	.0	0.99
$\pm$										0.15	0.09	.0	.0	0.17
2. Track resolution smearing =1.0														
ANN cut	1	215	114	79	34	12	8	10	9	32	14	0	0	
ϵ		.430	.228	.158	.068	.024	.016	.020	.018	.000143	.000933	.0	.0	
$\pm$		.022	.019	.016	.011	.007	.006	.006	.006	.000025	.000249	.0	.0	
$\langle n \rangle$										0.84	0.40	0.00	0.00	1.24
$\pm$										0.15	0.11	0.00	0.00	0.18
3. Track resolution smearing =0.9														
ANN cut	1	215	109	80	33	13	8	10	8	28	6	0	0	
ϵ		.430	.218	.160	.066	.026	.016	.020	.016	.000126	.000400	.0	.0	
$\pm$		.022	.018	.016	.011	.007	.006	.006	.006	.000024	.000163	.0	.0	
$\langle n \rangle$										0.73	0.17	.0	.0	0.91
$\pm$										0.14	0.07	.0	.0	0.16

Table 4.8: Track resolution and Si hit dropping systematic checks for the a63 network.

	data	$hA \to AAA$ (60,30)	(70,20)	(80,30)	(80,40)	$hZ \to ZAA$ (60,30)	(70,20)	(80,30)	(80,40)	$e^+e^- \to$ $q\bar{q}(g)$	$q\bar{q}q\bar{q}$	$q\bar{q}\ell^+\ell^-$	$q\bar{q}e^+e^-$	$\langle n \rangle$
1. Variation in Si hit matching efficiency														
ANN cut	1	85	38	214	218	46	47	79	62	15	16	0	0	
ϵ		.170	.076	.428	.436	.092	.094	.158	.124	.000075	.001067	.0	.0	
$\pm$		.017	.012	.022	.022	.013	.013	.016	.015	.000019	.000267	.0	.0	
$\langle n \rangle$										.44	.46	.0	.0	.90
$\pm$										.11	.11	.0	.0	.16
2. Track resolution smearing =1.0														
ANN cut	1	82	38	234	221	47	51	82	63	23	23	0	0	
ϵ		.164	.076	.468	.442	.094	.102	.164	.126	.000103	.001533	.0	.0	
$\pm$		.016	.012	.022	.022	.013	.014	.016	.015	.000021	.000319	.0	.0	
$\langle n \rangle$										.60	.66	.0	.0	1.26
$\pm$										.13	.14	.0	.0	.19
3. Track resolution smearing =0.9														
ANN cut	1	80	37	228	212	51	47	80	60	21	18	0	0	
ϵ		.160	.074	.456	.424	.102	.094	.160	.120	.000094	.001200	.0	.0	
$\pm$		.016	.012	.022	.022	.014	.013	.016	.014	.000021	.000283	.0	.0	
$\langle n \rangle$										.55	.51	.0	.0	1.07
$\pm$										.12	.12	.0	.0	.17

Table 4.9: Track resolution and Si hit dropping systematic checks for the a83 network.

4.6.3 Different Monte Carlo Generators

The background estimate is subject to possible mismodelling of the jet resolution parameters due to the complication of modelling higher-order QCD processes, including those in radiative events, and hadronization. The charged track multiplicity and electromagnetic cluster multiplicity distributions will be similarly affected by the jet evolution models chosen. In addition, the b-discriminate values discussed previously will be affected by differences in the simulation of B hadron formation and decays.

To estimate the sensitivity of the background estimate to errors related to the modelling of Standard Model physics, alternate Monte Carlo generators were used as a consistency check for the predicted background. This alternate background is described in Table 4.1. The result of running on this background sample is presented in Table 4.10 and is summarized in Table 4.11.

No alternate Monte Carlo is available for the signal events so the signal efficiencies are not changed by this systematic check and are not presented.

	data	$hA \to AAA$ (60,30)	(70,20)	(80,30)	(80,40)	$hZ \to ZAA$ (60,30)	(70,20)	(80,30)	(80,40)	$(Z/\gamma)^* \to f\bar{f}$	$W^+W^- \to f\bar{f}f\bar{f}$	$Z^0Z^{0(*)} \to f\bar{f}f\bar{f}$	$\langle n \rangle$
1. a63 network with the alternative background sample													
ANN cut	1	417	114	164	33	13	6	10	9	31	11	12	
ϵ		.417	.228	.164	.066	.026	.012	.020	.018	.000207	.000275	.002400	
$\pm$		.016	.019	.012	.011	.007	.005	.006	.006	.000037	.000083	.000692	
$\langle n \rangle$										1.21	0.23	0.08	1.52
$\pm$										0.22	0.07	0.02	0.23
1. a83 network with the alternative background sample													
ANN cut	1	154	37	459	222	48	47	84	66	17	16	23	
ϵ		.154	.074	.459	.444	.096	.094	.168	.132	.000113	.000400	.004600	
$\pm$		.011	.012	.016	.022	.013	.013	.017	.015	.000027	.000100	.000957	
$\langle n \rangle$										0.66	0.34	0.15	1.16
$\pm$										0.16	0.08	0.03	0.18

Table 4.10: Variation in the expected background $\langle n \rangle$ due to alternative Monte Carlo generators.

ANN Analysis	$\langle n \rangle (\Delta)$
a63	1.52(0.51)
a83	1.16(0.15)

Table 4.11: Variations in expected background ($\langle n \rangle$) is presented for the two neural networks in this analysis. The bracketed values then give the difference between a particular value compared to the primary result.

Not only is the difference of 0.51 of an event between the primary and alternative background samples for the a63 neural network the largest source of systematic error for that analysis, but this large difference was not observed in the a83 analysis.

Comparing Table 4.10 to the primary results in Table 4.14 will show that the expected contribution to the total background from the four fermion processes is consistent with the primary background sample within statistical errors. The source of the discrepancy is from the two fermion channels, simulated by PYTHIA in the primary analysis and HERWIG in the alternative sample. The fact that this discrepancy is not observed in the a83 analysis implies that the variables which are causing the difference are emphasized by the a63 neural net more than in the a83 neural net.

The source was isolated to a systematic difference between the predicted y^D_{45} and y^D_{56} distributions as shown in Figure 4.11. For the two $y^D_{n,n+1}$ variables presented, the left plot, (a,c), compares data (crossed points) to the main background (dashed) and the alternative background (dotted). Both distributions seem to model the data well, although the alternative background appears shifted to the right, which is in the direction of the $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV signal (dot-dash). The right hand plot, (b,d), shows the PYTHIA two fermion (negative slope hatch) component of the main background (dashed), compared with the HERWIG two fermion (positive slope hatch) of the alternative background (dotted). Unfortunately the statistics available are insufficient to make a statement about which Monte Carlo is most correct.

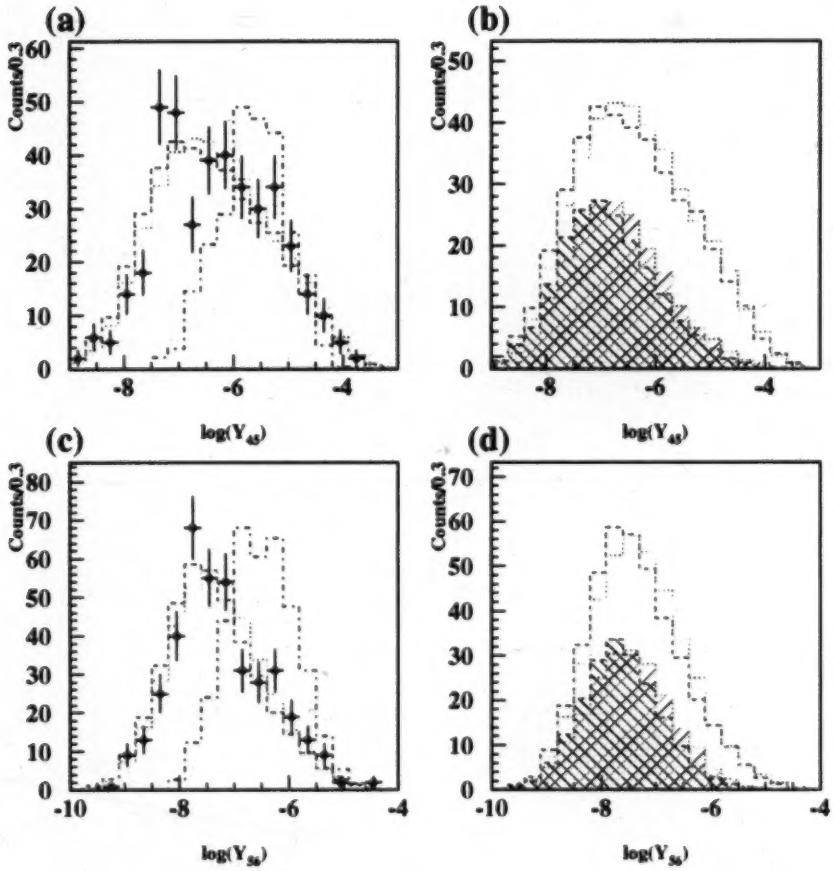


Figure 4.11: For the two variables shown, the left plot, (a,c), compares data (crossed points) to the main background (dashed) and the alternative background (dotted). The alternative looks more shifted to the right, which is in the direction of the (m_{h^0} , m_{A^0})=(60,30)GeVsignal (dot-dash). The right hand plots, (b,d), shows the PYTHIA two fermion (negative slope hatch) of the main background (dashed), compared with the HERWIG two fermion (positive slope hatch) of the alternative background (dotted).

4.6.4 ANN Systematic Error

In order to apply and interpret the Monte Carlo Transformation analysis, Section 3.5, one must check that the data distribution is consistent with a background only hypothesis, rather than a signal plus background hypothesis. In the latter case, a background subtraction method should be used first, which would make an interpretation of this method as a estimator of the total systematic error less clear.

In this analysis we see no significant excess in the data (cf., Table 4.13, Figure 4.14), which leads us to believe that we may proceed with the transformations as a total systematic error study. Figures 4.12, 4.13 show the effect of this Monte Carlo transformation technique, using $\mathcal{S}_{\text{ndf}} = 1.0$, applied to the ANN input distributions.

Given the discussion in Section 4.6.3, we pay special attention to the transformed $\log(y_{45}^D)$ and $\log(y_{56}^D)$ distributions. In both cases, the transformed distribution are shifted to the right which would imply that the HERWIG predicted distributions, which are right shifted relative to the PYTHIA distributions as discussed above, better agree with data.

Table 4.12 shows the results of applying the transformations to the inputs of the a63 and a83 networks. Rows 1 and 4 show that the effect of transforming all the inputs, and results in efficiency and expected background variations, consistent with the largest systematic error sources considered so far. To test whether the variations in $\log(y_{45}^D)$ and $\log(y_{56}^D)$ account for the majority of the discrepancy in the a63 expected background, rows 2 and 3 present the results for the a63, where the transformations were applied to the $\log(y_{45}^D), \log(y_{56}^D)$ and the remaining distributions separately. This seems to imply that $\log(y_{45}^D)$ and $\log(y_{56}^D)$ do in fact account for the observed differences, but this should not be taken too quantitatively since by transfroming these parameters alone, leaving the remaining inputs untouched, we are introducing a variation in the correlations which the neural network was trained with. This convolution is not easily removed.

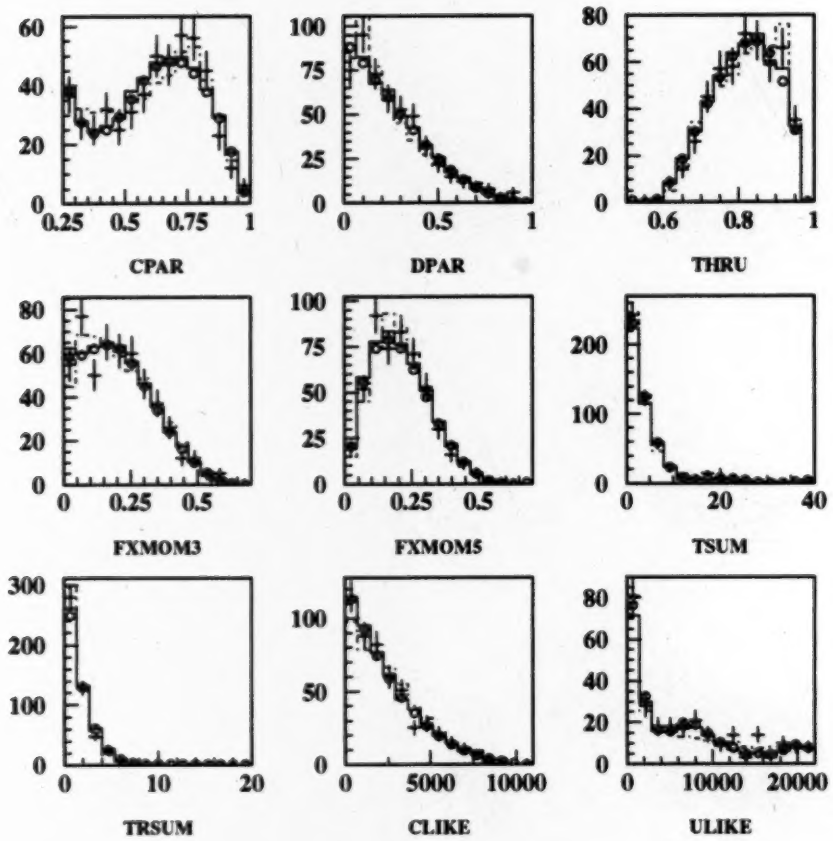


Figure 4.12: Input variables to the ANN after the Monte Carlo Transformation technique of Section 3.5 is applied. The spline fit (open circles) to the raw data distribution (crossed points) is the target for the raw Monte Carlo Background (dashed histogram). The Solid histogram is the transformed Monte Carlo distribution.

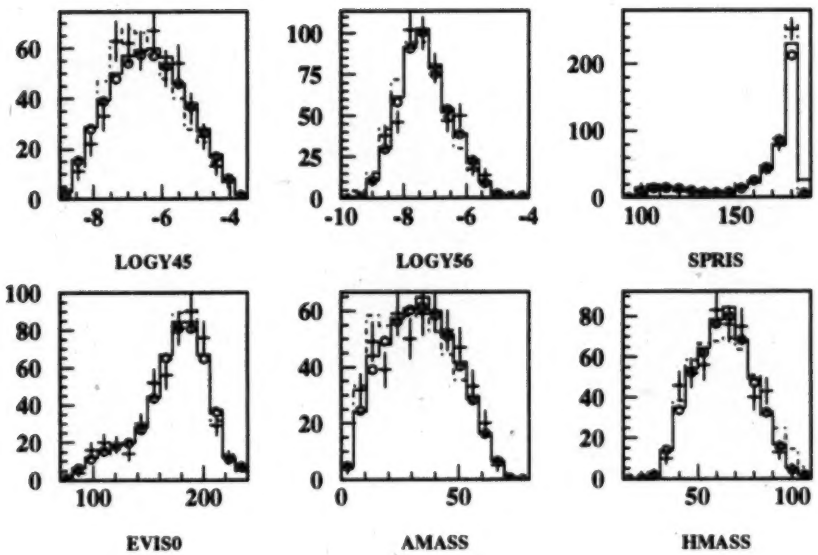


Figure 4.13: Input variables to the ANN after the Monte Carlo Transformation technique of Section 3.5 is applied. The spline fit (open circles) to the raw data distribution (crossed points) is the target for the raw Monte Carlo Background (dashed histogram). The Solid histogram is the transformed Monte Carlo distribution.

	data	$hA \rightarrow AAA$				$hZ \rightarrow ZAA$				$e^+e^- \rightarrow$				$\langle n \rangle$
		(60,30)	(70,20)	(80,30)	(80,40)	(60,30)	(70,20)	(80,30)	(80,40)	$q\bar{q}(g)$	$q\bar{q}q\bar{q}$	$q\bar{q}\ell^+\ell^-$	$q\bar{q}e^+e^-$	
1. a63 ANN where *all* inputs have been transformed														
ANN cut	1	440	113	156	30	15	8	9	8	40	6	0	0	
ϵ		.440	.226	.156	.060	.030	.016	.018	.016	.000178	.000400	.0	.0	
$\pm$		.016	.019	.011	.012	.006	.006	.006	.006	.000028	.000163	.0	.0	
$\langle n \rangle$										1.04	.17	.0	.0	1.21
$\pm$										.16	.07	.0	.0	.18
2. a63 ANN where all input *excluding* y-parameters are transformed														
ANN cut	1	414	104	142	29	13	5	9	7	32	6	0	0	
ϵ		.414	.208	.142	.058	.026	.010	.018	.014	.000142	.000400	.0	.0	
$\pm$		.016	.018	.011	.010	.007	.004	.006	.005	.000025	.000163	.0	.0	
$\langle n \rangle$										0.83	0.17	.0	.0	1.00
$\pm$										0.15	0.07	.0	.0	0.16
2. a63 ANN where *only* y-parameters are transformed														
ANN cut	1	449	126	172	33	15	8	10	10	42	11	0	0	
ϵ		.449	.252	.172	.066	.030	.016	.020	.020	.000187	.000733	.0	.0	
$\pm$		.017	.019	.012	.011	.008	.006	.006	.006	.000029	.000221	.0	.0	
$\langle n \rangle$										1.09	0.31	0.00	0.00	1.41
$\pm$										0.17	0.09	0.00	0.00	0.19
1. a83 ANN where *all* inputs have been transformed														
ANN cut	1.	237.	65.	470.	219.	46.	49.	83.	64.	24.	22.	0.	0.	
ϵ		.237	.130	.470	.438	.092	.098	.166	.128	.000107	.001467	.0	.0	
$\pm$		.013	.015	.016	.022	.013	.013	.017	.015	.000022	.000312	.0	.0	
$\langle n \rangle$										0.62	0.63	.0	.0	1.25
$\pm$										0.13	0.13	.0	.0	0.18

Table 4.12: Results of the Monte Carlo Transformation method are presented. Tables 1-3 are for the a63 neural network, Table 4 is for the a83 neural network. Compare to Table 4.13.

4.7 Results

Figure 4.14 shows the neural network distributions for the 400 data events which passed the preselection in addition to the expected background normalized to the observed luminosity and the associated signal distribution normalized to the number of observed data.

Summarizing then, the analysis applied was:

1. Preselection cuts:
   - The LEP Multihadronic event selection flag was set.
   - The polar angle of the thrust axis, θ_T, is required to satisfy $|\cos\theta_T| \leq 0.9$.
   - Charged track multiplicity ≥ 25.
   - The C-parameter is required to be ≥ 0.25.
   - The effective centre of mass energy, $\sqrt{s} \geq 100\,\text{GeV}$.
   - 6 jets were reconstructed.
2. Jet quality cuts. Jet multiplicity ≥ 2 and EM cluster multiplicity ≥ 1.
3. The event 4-C fit was required to succeed.
4. Events passed the appropriate neural network cut.

Table 4.13 presents the results for the two neural networks, a63 and a83, applied to the main background sample as well as the $h^0A^0 \to A^0A^0A^0$ and $h^0Z^0 \to Z^0A^0A^0$ at various mass points. Cuts $1-3$ are common to both neural networks. The fourth cut in this Table, as well as in the remaining Tables in this section, is divided into a63 and a83 analysis results. The integral values along the first row of each cut section are the actual number of events from the sample passing the cuts. The ϵ is the efficiency and $\langle n \rangle$ is the expected number of events, here at $\mathcal{L} = 54.5 \pm 0.2\ \text{pb}^{-1}$. The extra column is the total sum of the expected background. Errors are statistical only. Table 4.14 summarizes the systematic errors, which are consistent with the Monte Carlo Transformation method predictions.

With a neural network output cut of ≥ 0.95, the $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV targeted analysis predicts $1.01 \pm 0.16 \pm 0.56$ event background at $\mathcal{L} = 54.5 \pm 0.2\ \text{pb}^{-1}$ with a (60, 30) GeV signal efficiency $0.417 \pm 0.016 \pm 0.051$, where the errors shown are statistical and systematic, respectively. One event was observed in the data (`Run=8294, Event=12343`), which is consistent with the expected background. This event, shown in Figure 4.15, is compared to the SM predicted distributions in Figures 4.16 - 4.17.

With a neural network output cut of ≥ 0.965, the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV targeted analysis predicts $1.01 \pm 0.17 \pm 0.31$ event background at $\mathcal{L} = 54.5 \pm 0.2\ \text{pb}^{-1}$ with signal efficiencies $\epsilon_{(80,30)} = 0.459 \pm 0.016 \pm 0.071$ and $\epsilon_{(80,40)} = 0.444 \pm 0.022 \pm 0.049$, where $\epsilon_{(m_{h^0}, m_{A^0})}$ is the efficiency at the mass point (m_{h^0}, m_{A^0}), where the errors shown are statistical and systematic, respectively. One event was observed in the data (`Run=8712, Event=89200`), which is consistent with the expected background. This event, shown in Figure 4.18, is compared to the SM predicted distributions in Figures 4.19 - 4.20.

	data	$hA \to AAA$				$hZ \to ZAA$				$e^+e^- \to$				$\langle n \rangle$
		(60,30)	(70,20)	(80,30)	(80,40)	(60,30)	(70,20)	(80,30)	(80,40)	$q\bar{q}(g)$	$q\bar{q}q\bar{q}$	$q\bar{q}\ell^+\ell^-$	$q\bar{q}e^+e^-$	
Total	6131	1000	500	1000	500	500	500	500	500	224999	15000	40566	100000	
σ										107.43	7.86	8.11	26.73	
$\langle n \rangle$										5853.5	428.4	442.0	1456.7	8180.7
Preselection	846	921	467	970	476	431	442	428	439	20522	11445	3762	683	
ϵ		.921	.934	.970	.952	.862	.884	.856	.878	.0912	.7630	.0927	.0068	
$\langle n \rangle$										533.90	326.88	40.99	9.95	911.72
Jet quality	460	763	357	816	430	364	362	369	383	10547	8005	740	167	
ϵ		.763	.714	.816	.860	.728	.724	.738	.766	.0468	.5336	.0182	.0017	
$\langle n \rangle$										274.39	228.63	8.06	2.43	513.51
Event 4C fit	400	703	342	802	425	360	355	368	377	8342	7901	423	63	
ϵ		.703	.684	.802	.850	.720	.710	.736	.754	.0371	.5267	.0104	.0006	
$\langle n \rangle$										217.02	225.66	4.61	0.92	448.21
2. a63 trained network														
ANN cut	1	417	114	164	33	13	6	10	9	29	9	0	0	
ϵ		.417	.228	.164	.066	.026	.012	.020	.018	.000129	.000600	.0	.0	
$\pm$		.016	.019	.012	.011	.007	.005	.006	.006	.000024	.000200	.0	.0	
$\langle n \rangle$										0.75	0.26	.0	.0	1.01
$\pm$										0.14	0.09	.0	.0	0.16
3. a83 trained network														
ANN cut	1	154	37	459	222	48	47	84	66	19	18	0	0	
ϵ		.154	.074	.459	.444	.096	.094	.168	.132	.000084	.001200	.0	.0	
$\pm$		.011	.012	.016	.022	.013	.013	.017	.015	.000019	.000283	.0	.0	
$\langle n \rangle$										.49	.51	.00	.00	1.01
$\pm$										.11	.12	.00	.00	.17

Table 4.13: Results for the main background sample. Integral values along the first row in each cut section are the actual number of events from each sample passing the cuts. ϵ is the efficiency and $\langle n \rangle$ is the expected number of events, here at $\mathcal{L} = 54.5 \pm 0.2$ pb^{-1}. The extra column is the total sum of the expected background. In the 'Total' row, σ is the cross-section of the process and $\langle n \rangle$ is $\sigma\mathcal{L}$. The preselection, jet quality and 4-C fit cuts were common to both networks and the application of the trained networks are presented seperately in sub-tables 2 and 3. Errors are statistical only.

Check	$hA \rightarrow AAA$, $\Delta\%$				$hZ \rightarrow ZAA$, $\Delta\%$				$\langle n \rangle$, $\Delta\%$
	(60,30)	(70,20)	(80,30)	(80,40)	(60,30)	(70,20)	(80,30)	(80,40)	
1. a63 trained network									
Preselection	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Si Hit Dropping	-4.077	2.632	4.878	3.030	.000	16.667	.000	-11.111	1.980
Si Resolution=1.1	-3.118	.000	3.659	-3.030	7.692	-33.333	.000	.000	-22.772
Si Resolution=0.9	-3.118	4.386	2.439	.000	.000	-33.333	.000	11.111	9.901
MC Generators	-	-	-	-	-	-	-	-	-50.49
Total Systematic	±5.13	±5.11	±6.10	±4.28	±7.69	±37.26	±0.00	±11.11	±55.42
2. a83 trained network									
Preselection	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Si Hit Dropping	-10.390	-2.703	6.754	1.802	4.167	.000	5.952	6.061	10.891
Si Resolution=1.1	-6.494	-2.703	-1.961	.450	2.083	-8.511	2.381	4.545	-24.752
Si Resolution=0.9	-3.896	.000	.654	4.505	-6.250	.000	4.762	9.091	-5.941
MC Generators	-	-	-	-	-	-	-	-	-14.851
Total Systematic	±12.25	±3.82	±7.06	±4.87	±7.51	±8.51	±7.62	±10.92	±30.85

Table 4.14: This table summarises the systematic uncertainties determined for the signal efficiencies and the expected background in both the analysis. These uncertainities are then summed in quadrature, taking the maximum variation from the Si Resolution rows, to give the total uncertainty per channel.

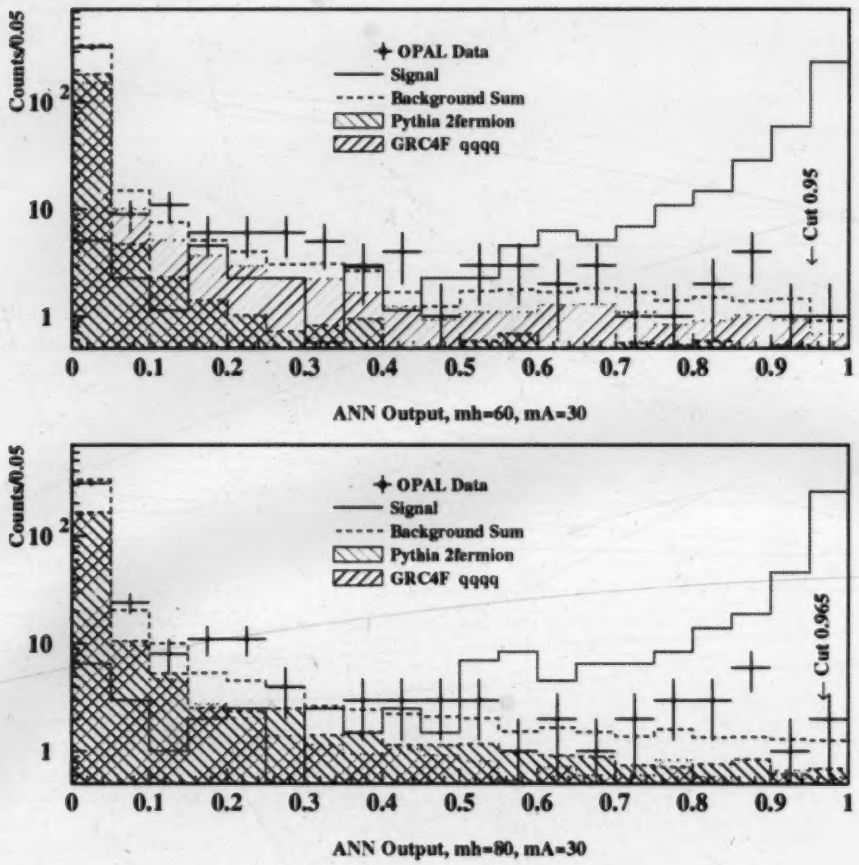


Figure 4.14: Presented are the neural network output distributions for the two trained networks, the (60,30) and (80,30). The signal distributions are for the mass values at which the network was trained.

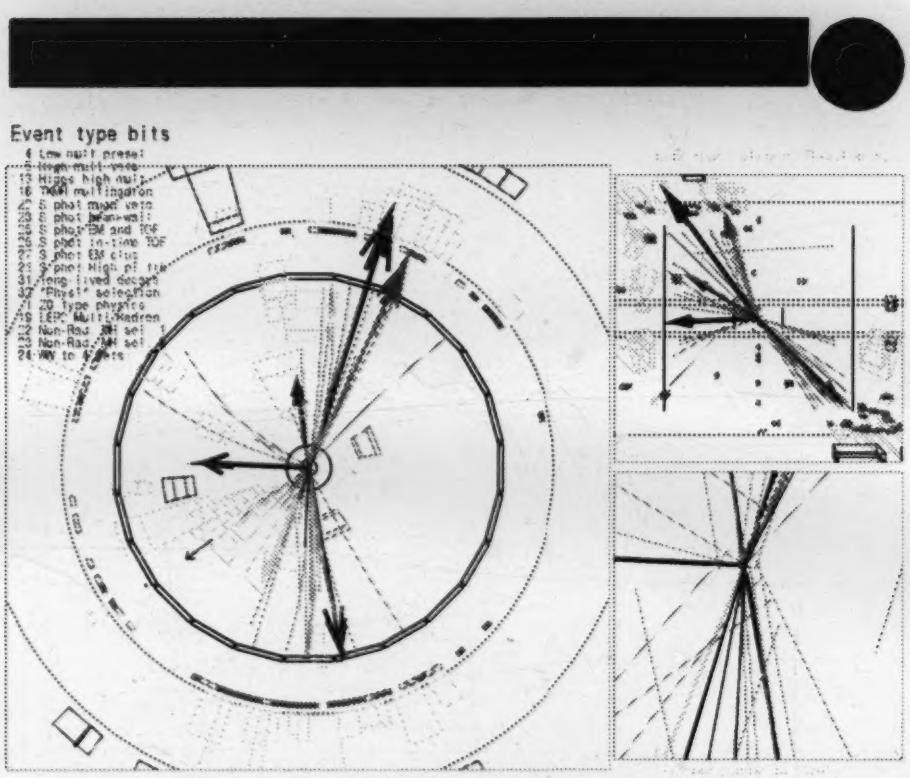


Figure 4.15: Event display of event 12343 in run 8294, selected by the a63 neural network in the $\sqrt{s} = 183$ GeV analysis as a signal candidate.

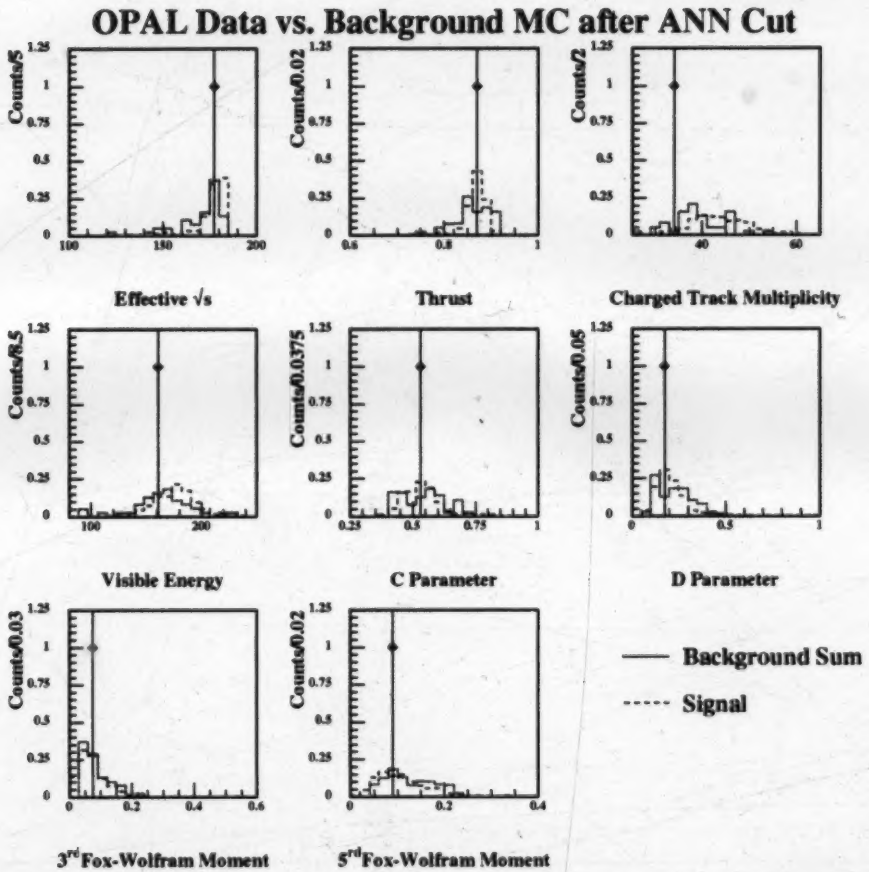


Figure 4.16: Distribution of the Neural Network input parameters for OPAL data (full circles), (60,30) signal (dashed) and the SM prediction (open histogram) after application of the a63 network. The data point surviving is consistent with the SM prediction.

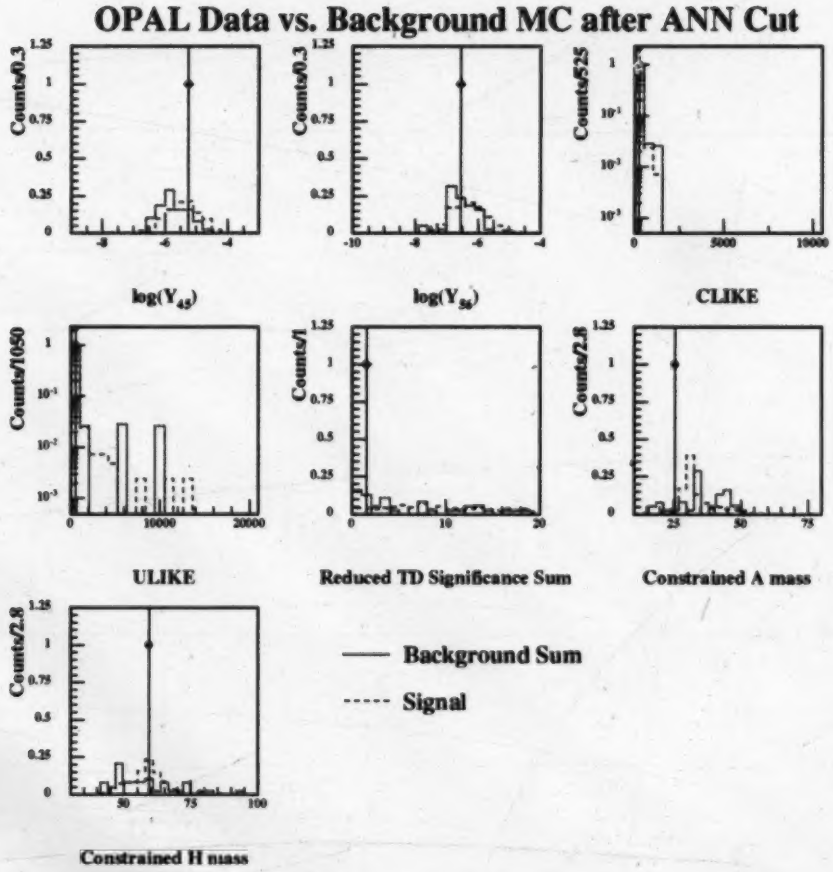


Figure 4.17: Distribution of the Neural Network input parameters for OPAL data (full circles), (60,30) signal (dashed) and the SM prediction (open histogram) after application of the a63 network. The data point surviving is consistent with the SM prediction.

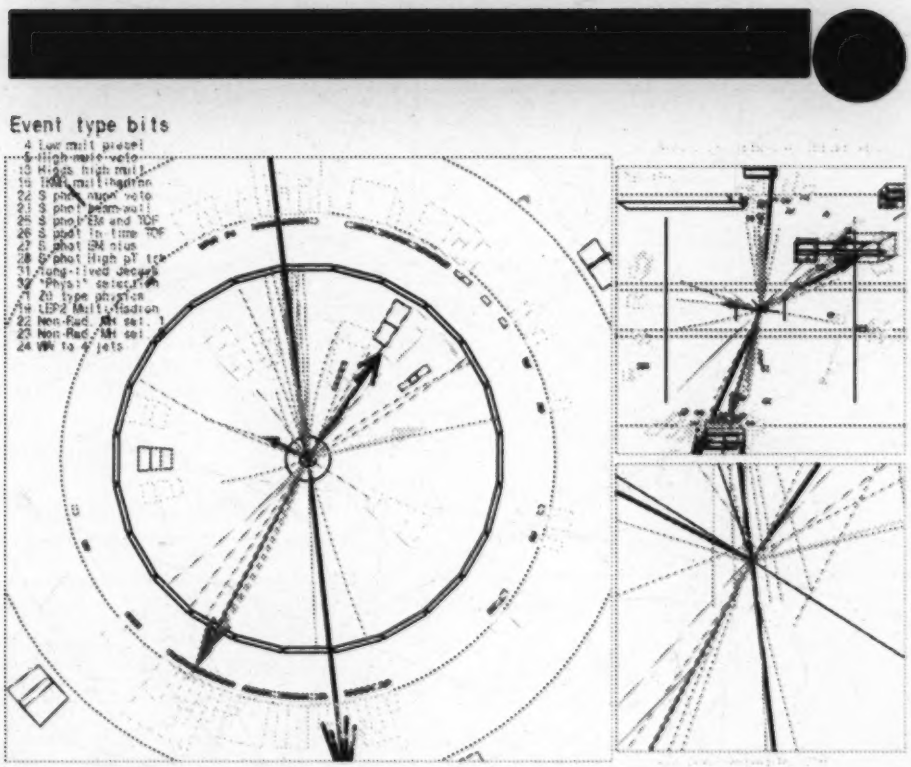


Figure 4.18: Event display of event 89200 in run 8712, selected by the a83 neural network in the $\sqrt{s} = 183$ GeV analysis as a signal candidate.

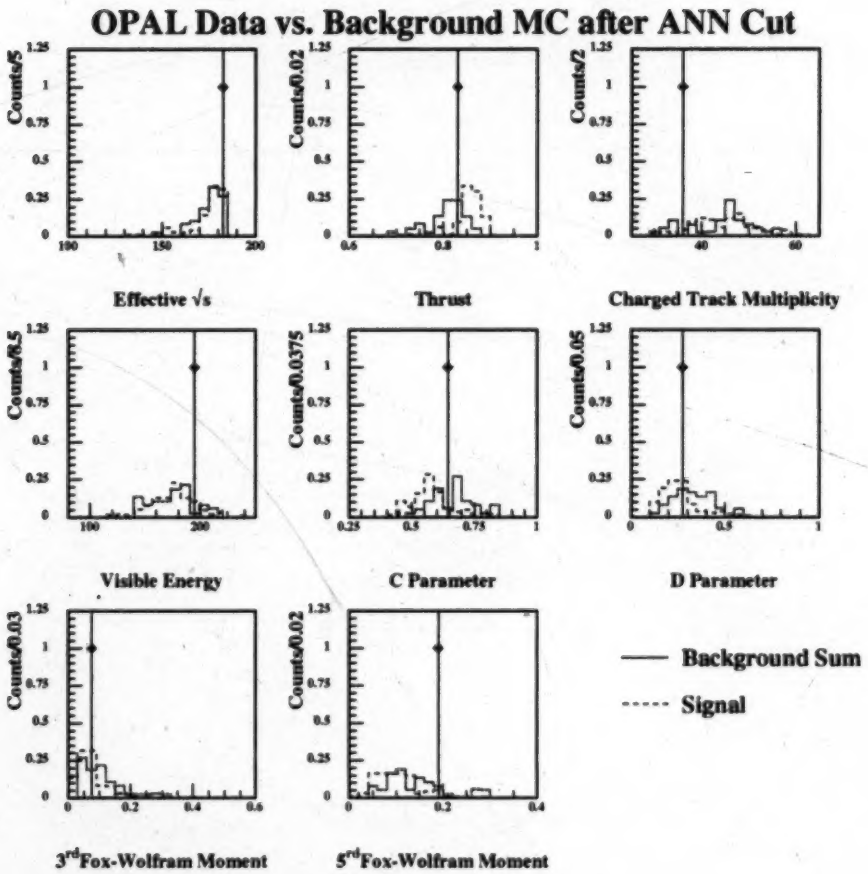


Figure 4.19: Distribution of the Neural Network input parameters for OPAL data (full circles), (80,30) signal (dashed) and the SM prediction (open histogram) after application of the a83 network. The data point surviving is consistent with the SM prediction.

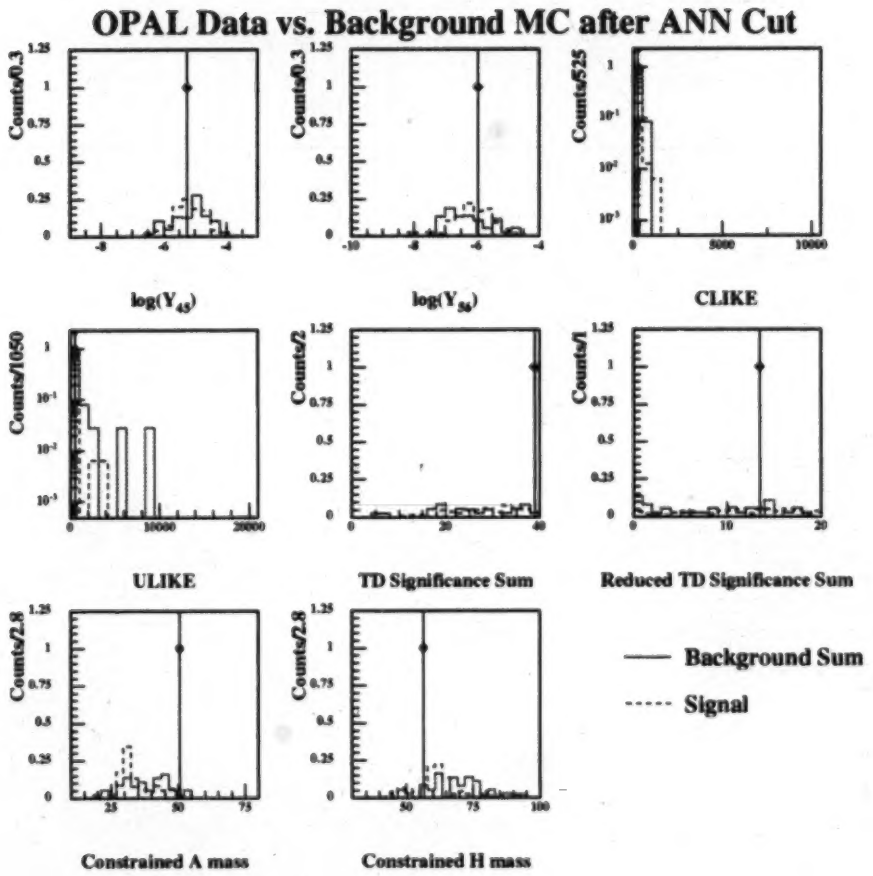


Figure 4.20: Distribution of the Neural Network input parameters for OPAL data (full circles), (80,30) signal (dashed) and the SM prediction (open histogram) after application of the a83 network. The data point surviving is consistent with the SM prediction.

4.7.1 MSSM Exclusion Limit

As discussed in section 1.2.2, a broad range of MSSM model parameters (listed on page 14) which affect the production of signal events must be considered in order to place exclusion limits for all possible values of these parameters. A constrained MSSM was considered when deriving the exclusion limits. This model is parameterized by six free parameters in addition to those of the SM: unification of the scalar-fermion masses (m_0) at the grand unification (GUT) scale; unification of the gaugino masses (which are parametrized using M_2, the SU(2) gaugino mass term at the electroweak scale); and scalar-fermion tri-linear couplings (A) at the electroweak scale. Other free parameters of the model are the supersymmetric Higgs mass parameter μ, $\tan\beta$, and the mass of the CP-odd neutral Higgs boson, m_{A^0}. The top quark mass has a strong impact on m_{h^0}; therefore, it was also varied within reasonable bounds.

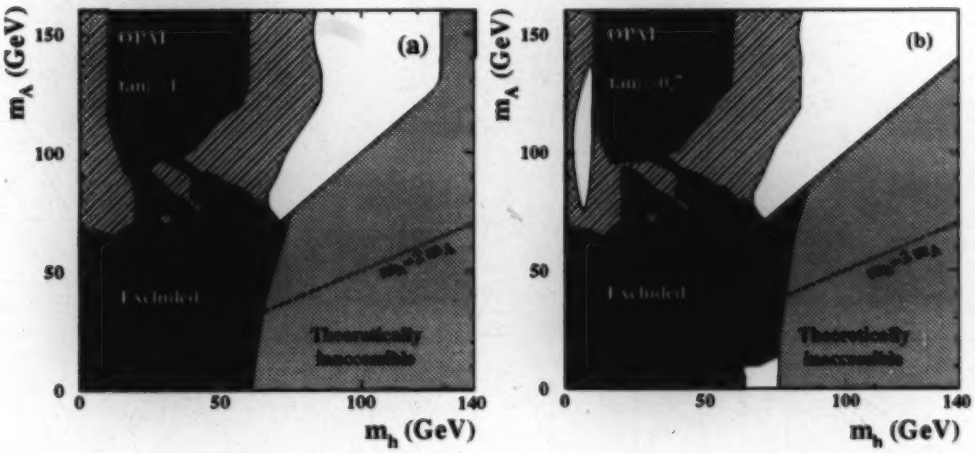


Figure 4.21: The MSSM exclusion for the 183 GeV analysis. Excluded regions are shown for (a) the (m_{h^0}, m_{A^0}) plane for $\tan\beta > 1$, (b) the (m_{h^0}, m_{A^0}) plane for $\tan\beta > 0.7$. All exclusion limits are at 95% CL. The black areas are excluded solely by comparing data to the expected MSSM production rates, without applying any additional theoretical criteria. The non-hatched grey areas are theoretically inaccessible.

The exclusion limits from the $\sqrt{s} = 172\,\text{GeV}$ analysis were combined with the results from this new scan of the MSSM parameter space, which used results from all OPAL Higgs analyses performed at $\sqrt{s} = 183\,\text{GeV}$ [104], to create the 95% CL exclusion limits shown in Figure 4.21. The black areas are excluded solely by comparing data to the expected MSSM production rates, without applying any additional theoretical criteria. The non-hatched grey areas are theoretically inaccessible.

For $\tan\beta > 1.0$ (Figure 4.21a), the regions where $2m_{A^0} \leq m_{h^0}$, which were left unexcluded by the previous $\sqrt{s} = 172\,\text{GeV}$, are now excluded. When the $\tan\beta$ range is enlarged to include $\tan\beta > 0.7$ (Figure 4.21b), the exclusion region is now seen to cover the entire theoretically accessible area up to $m_{h^0} \approx 75\,\text{GeV}$, except for a small unexcluded region at 5 GeV $< m_{A^0} <$ 15 GeV and 65 GeV $< m_{h^0} <$ 72 GeV.

The next chapter will focus on closing this remaining unexcluded region at a higher centre of mass energy.

Chapter 5

The Search for $e^+e^- \to Z^0 \to h^0A^0 \to A^0A^0A^0 \to 3(b\bar{b})$ at $\sqrt{s} = 189\,\mathrm{GeV}$

In this chapter, we continue the search for Higgs production through channel $e^+e^- \to h^0\,A^0 \to A^0\,A^0\,A^0 \to 3(b\bar{b})$, at an increased centre of mass energy of 189 GeV.

Whereas the 183 GeV analysis presented in the last chapter focussed on the highest accessible mass region, $m_{Z^0} \leq m_{h^0} + m_{A^0} \leq \sqrt{s}, m_{h^0} \simeq 2m_{A^0}$, that analysis identified a region, around $(m_{h^0}, m_{A^0}) \approx (60, 12)$ GeV, which was not excluded in the MSSM. It was decided, then, to expand the analysis with the aim of achieving high signal efficiency with low expected background over the entire signal region $m_{h^0} \geq 2m_{A^0}$, with the constraint that $m_{A^0} > 2m_b$, where m_b is the b-quark mass, given by the Particle Data Group [80] as $m_b = 4.1 - 4.5$ GeV. This would give the minimal m_{A^0} value as $m_{A^0} \approx 10$ GeV. However, this sits almost directly upon a $b\bar{b}$ resonance, Υ, so the minimum m_{A^0} was taken to be 12 GeV.

Note that the relevant (m_{h^0}, m_{A^0}) range can be split into two regions: $m_{h^0} + m_{A^0} > m_{Z^0}$ and $m_{h^0} + m_{A^0} < m_{Z^0}$. In the latter region, radiative return events are quite common and no cut on $\sqrt{s'}$ is possible without sacrificing signal efficiency.

Further, depending on the (m_{h^0}, m_{A^0}) mass point and whether radiative return occurred, many of the Monte Carlo signal events could be better described as three jets, where the $A^0 \to b\bar{b}$ decays are sufficiently boosted so as to look like a single jet. At higher (m_{h^0}, m_{A^0}) mass points the three $A^0 \to b\bar{b}$ decays can be well separated and identified.

The analysis presented here is based on data collected by the OPAL detector during the 1998-1999 run, representing a luminosity of $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\ \mathrm{pb}^{-1}$. The selection of candidates was performed in two ways: first, a simple cutting analysis, which emphasizes heavy flavour tagging, was used in order to gauge the expected background level; a more detailed Neural Network based analysis was then performed and found to yield higher signal efficiencies for the same expected background.

5.1 Data and Monte Carlo

The signal processes, $e^+e^- \to h^0\,A^0 \to 3\,A^0 \to 3(b\bar{b})$ with various m_{h^0}, m_{A^0}, were generated at 189 GeV using `HZHA2.00` [29] with full simulation of initial state radiation. The main contributions to the expected background are from those processes which have high charged track multiplicity and which can mimic multi-jet final states. The most important backgrounds contributing to this analysis were composed of the two fermion processes (with initial state radiation), $(Z/\gamma)^* \to f\bar{f}$

generated by PYTHIA [83], and the four fermion final states from W^+W^- and $Z^0Z^{0(*)}$ generated using GRC4F [84]. For systematic checks, described more fully below, an alternative Standard Model background sample was also generated using HERWIG [85] to simulate the $(Z/\gamma)^* \to q\bar{q}$ contribution and PYTHIA for the remaining allowed W^+W^- and $Z^0Z^{0(*)}$ channels. Events were then processed through a full simulation of the OPAL detector using the GOR021 [86] module with the OPAL98 detector constants and calibrations.

The signal and Standard Model background samples used to optimize the selection cuts are listed in Table 5.1.

Process	File	Number of Events
Data, representing $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50$ pb^{-1}		
Data	Period 91, Pass 6	1352
Data	Period 92, Pass 6	3224
Data	Period 93, Pass 6	3569
Data	Period 94, Pass 6	3809
Data	Period 95, Pass 6	2153
Data	Period 96, Pass 6	3969
HZHA generated signal, $Zh \to 3(b\bar{b})$		
$m_{h^0} = 40$, $m_{A^0} = 12$	...L20012/P189HA/R8444	2.5K
$m_{h^0} = 40$, $m_{A^0} = 20$	...L20012/P189HA/R8445	2.5K
$m_{h^0} = 50$, $m_{A^0} = 20$	...L20012/P189HA/R8446	2.5K
$m_{h^0} = 60$, $m_{A^0} = 12$	...L20012/P189HA/R8447	2.5K
$m_{h^0} = 60$, $m_{A^0} = 30$	...L20012/P189HA/R8448	2.5K
$m_{h^0} = 70$, $m_{A^0} = 20$	...L20012/P189HA/R8449	2.5K
$m_{h^0} = 80$, $m_{A^0} = 12$	...L20012/P189HA/R8450	2.5K
$m_{h^0} = 80$, $m_{A^0} = 30$	...L20012/P189HA/R8451	2.5K
$m_{h^0} = 80$, $m_{A^0} = 40$	...L20012/P189HA/R7873/P0010	0.5K
Main PYTHIA, GRC4F background		
$(Z/\gamma)^* \to q\bar{q}$	...L20012/P189QQ/R5078	225K
$e^+e^- \to q\bar{q}q\bar{q}$	...L20012/GRC4F/R7846	20K
$e^+e^- \to qeqe$	...L20012/GRC4F/R7849	30K
$e^+e^- \to q\ell q\ell$	...L20012/GRC4F/R8055	20K
Alternative HERWIG, PYTHIA background		
$(Z/\gamma)^* \to q\bar{q}$	...PASS12/HW59MH/R5080	100K
$W^+W^- \to$ all	...L20012/P189WW/R8027	20K
$Z^0Z^{0(*)} \to$ all	...L20012/P189ZZ/R8028	20K

Table 5.1: The signal and Standard Model background samples used in the 189 GeV analysis.

5.2 Preselection

In order to create manageable data sets, certain preselections were applied at the event reconstruction level.

5.2.1 Subdetector Status

The following subdetectors were required to have good status flags (cf. Table 4.2): SI, CV, CJ, EB, EE, HT, as well as SW (to ensure a good luminosity calculation). Exceptionally, due a non-critical HP gas problem in period 96 the status flag for HP was checked to be ≥ 2, to prevent the loss of data from this period.

Using this detector status selection, the total luminosity for the data set analysed, calculated using ROCROS [88], was found to be $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50$ pb^{-1}.

5.2.2 Event Reconstruction

The analysis variables were generated using OPAL Higgs group's analysis package, NT [112], generalized to n-jets by the author.

Tracks, electromagnetic and hadronic clusters were identified using the same selection criteria as in the 183 GeV analysis (Section 4.3.3). Double counting of the energy deposition of tracks is accounted for using the MT [92] package. The corrected tracks and clusters were then used to reconstruct a six jet topology using the Durham [96] algorithm. Events which failed to be reconstructed as six jets were rejected.

5.2.3 Changes from 183 GeV analysis

As the focus of this analysis shifted from the high mass regions considered in the 183 GeV analysis, to include lower mass regions, a number of the preselection cuts described in Section 4.3 had to be altered or removed altogether. Figures 5.1 to 5.4 show the comparison between data and monte carlo for the kinematic and topological variables considered in this analysis.

First, the mean number of charged tracks, shown in Figure 5.1, in the low mass region signal was lower than that seen in the high mass region of the 183 GeV analysis, so the preselection cut was reduced from 25 tracks to 20. In this, and the subsequent figures, two mass points are overlaid, $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 12)$ GeV satisfying $m_{\mathrm{h}^0} + m_{\mathrm{A}^0} < m_{\mathrm{Z}^0}$ and $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (80, 30)$ GeV satisfying $m_{\mathrm{h}^0} + m_{\mathrm{A}^0} > m_{\mathrm{Z}^0}$, from which we can see the change in distributions of the topological parameters over the allowed mass range.

One of the more powerful preselection cuts in the previous analysis was based on the effective centre of mass energy of the event, $\sqrt{s'}$. In the previous analysis, the signal points of interest had $\sqrt{s'} \approx \sqrt{s}$ and so the radiative return background events could be efficiently rejected with a minimum $\sqrt{s'}$ requirement. As can be seen in the $\sqrt{s'}$ distribution in Figure 5.1, signal points in the lower mass region show significant ISR production and so this cut was removed.

As well, the topology of the signal events now includes those events, in the region $m_{\mathrm{A}^0} \ll m_{\mathrm{h}^0}$, which are better classified as three jet rather than six jet events. Figure 5.4 presents the Durham *y-cut* distributions for various numbers of reconstructed jets. As we can see in this figure, the distribution of the lighter signal mass point, $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 12)$ GeV, is more similar to the two fermion component of the Standard Model background than is the distribution of the heavier signal mass point, $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (80, 30)$ GeV, which can be seen to be more distinct from the expected Standard Model background distribution, as we would expect from a true multi-jet event generated with little boost. In particular, the preselection efficiencies of the lighter signal points were detrimentally affected by the high C-parameter preselection cut in the 183 GeV analysis where it was applied to reduce the $(\mathrm{Z}/\gamma)^* \to \mathrm{f\bar{f}}$ background. Thus, the constraint on the C-parameter, seen in Figure 5.2, was lowered and is now required to be ≥ 0.0075.

To provide discrimination lost by the easing of the $\sqrt{s'}$ and C-parameter selection cuts, a new preselection cut, based on y_{43}^D, was introduced. y_{43}^D, the jet resolution parameter in the Durham scheme at which events change from 3 to 4 jets, was required to be ≥ 0.0005. The reason for the choice of y_{43}^D instead of, say y_{32}^D or y_{65}^D, was to bring this selection cut in line with the other multi-jet analysis in the Higgs group making comparisons of the selection efficiencies between the various search channels easier.

To recover the rejection power lost due to the relaxation of the kinematics preselection cuts, a loose b-tag was introduced. The jets were ordered in decreasing value of their b-discriminate, $\mathcal{B}_{\text{Jet}}$, which is related to their probability of being b-flavoured jets through Equation 5.1 described on page 131. Figure 5.7 presents the distributions of the first five of these ordered jet b-discriminates. From this ordered list, the jet with the largest b-discriminate is required to have $\mathcal{B}_{\text{Jet}} > 0.3$. The final results were found to be independent of this preselection cut, as can be seen in the distribution of this variable for those events passing the analysis (Figure 5.24).

None of these changes had a significant detrimental effect on the high mass points, $(m_{\text{h}^0}, m_{\text{A}^0}) = \{(60,30),(80,30)\}$ GeV, and only improved the preselection efficiencies for the low mass region, namely $m_{\text{A}^0} \ll m_{\text{h}^0}$ and $m_{\text{h}^0} + m_{\text{A}^0} \leq m_{\text{Z}^0}$.

5.2.4 Summary of Preselection Cuts

Summarizing then, the preselection applied was:

- The LEP Multihadronic event selection flag was set.
- The polar angle of the thrust axis, θ_T, is required to satisfy $|\cos\theta_T| \leq 0.9$ so that the event is well contained in the sensitive areas of the detector.
- Charged track multiplicity ≥ 20.
- The C-parameter is required to be ≥ 0.0075.
- $y_{43}^D \geq 0.0005$.
- 6 jets were reconstructed.
- Jet with the largest b-discriminate is required to have $\mathcal{B}_{\text{Jet}} > 0.3$.

The Monte Carlo signal survived these preselections with an efficiency > 0.95 and reduced the background by $\approx 88\%$.

Once manageable analysis data sets were created, the following selections were applied at analysis level,

- The jets' charged track and electromagnetic cluster multiplicity ≥ 2 and ≥ 1 respectively.

This last cut was not subject to optimization as the goal was only to ensure good event quality with well defined jets.

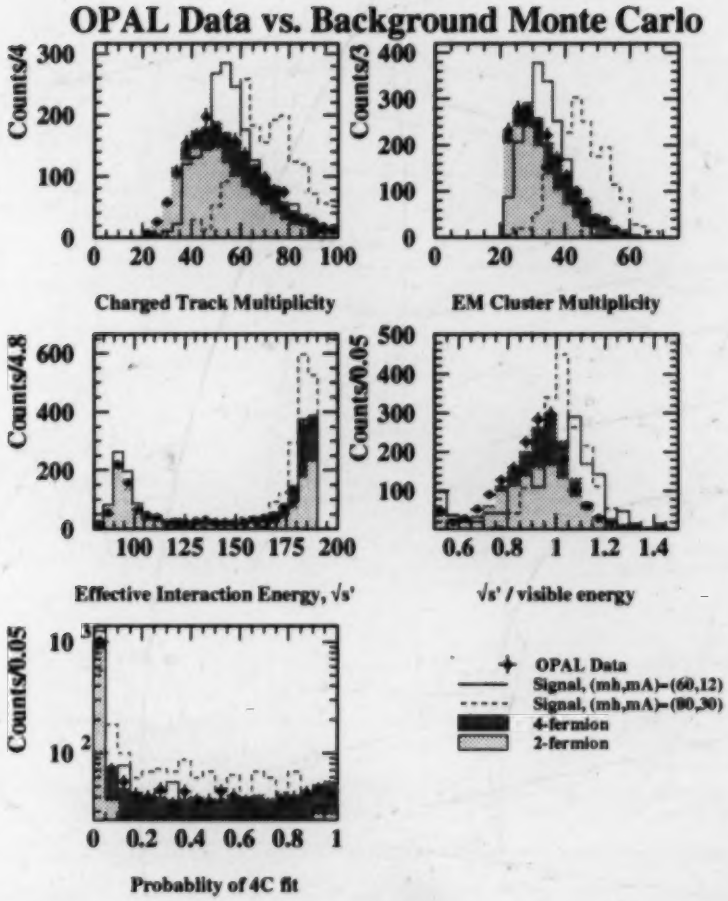


Figure 5.1: Distributions of the relevant kinematic analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

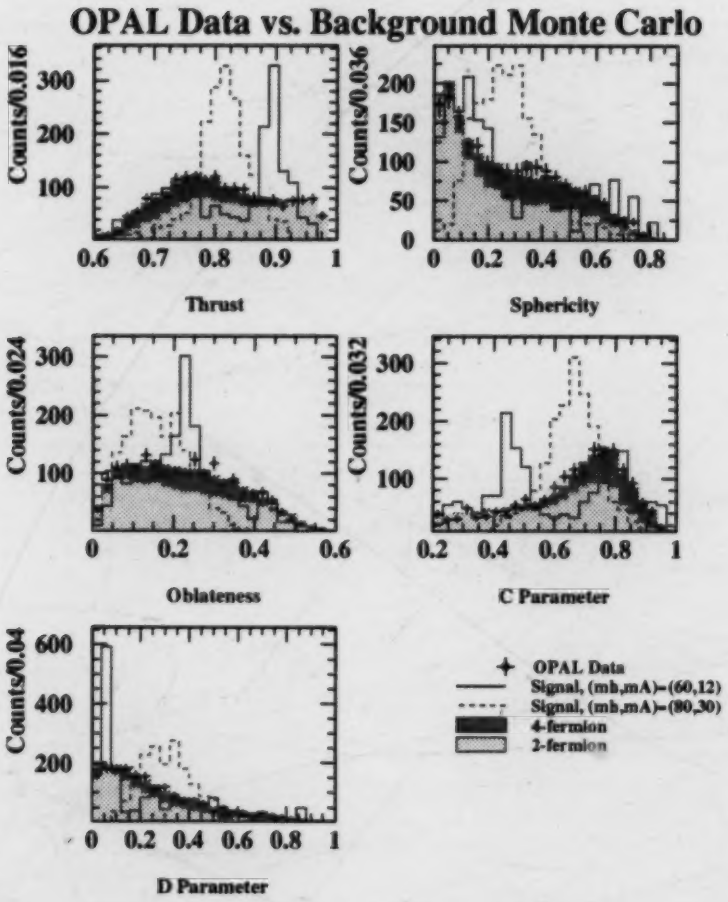


Figure 5.2: Distributions of the relevant kinematic analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

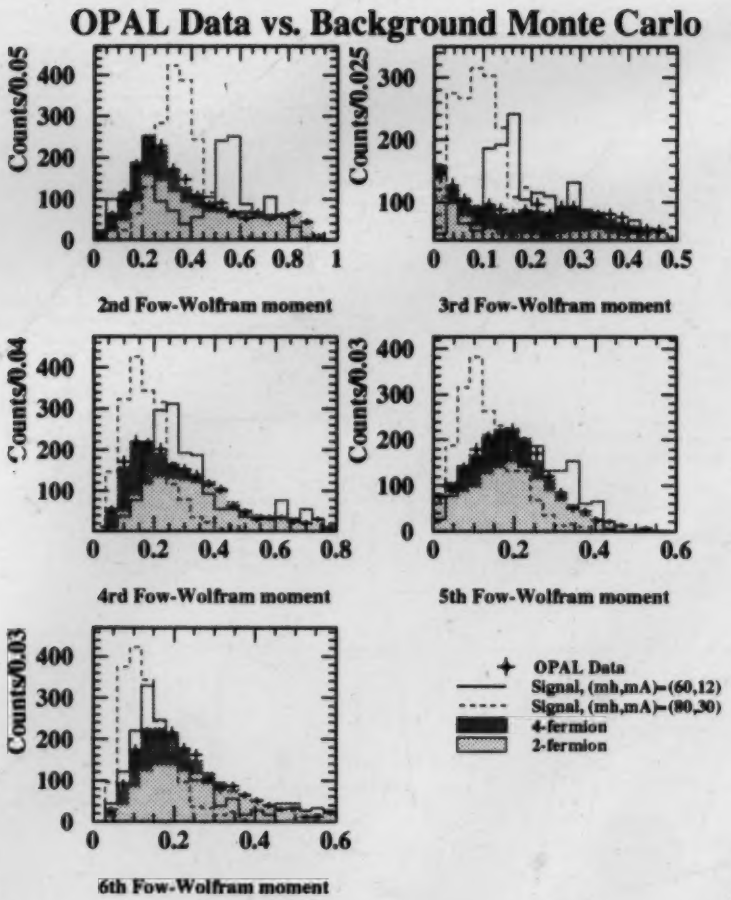


Figure 5.3: Distributions of the relevant kinematic analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

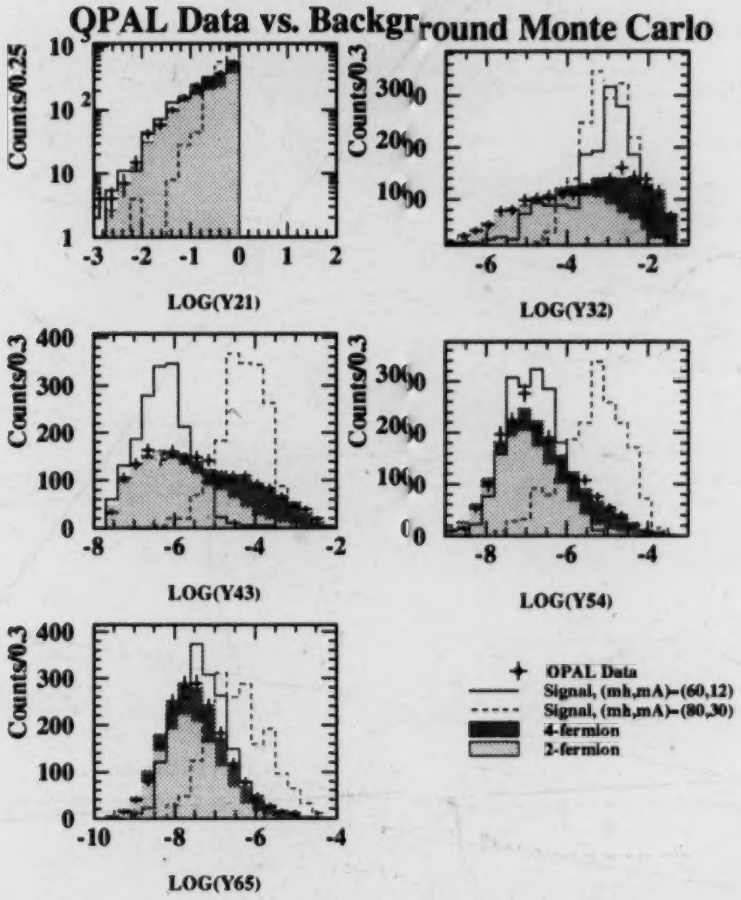


Figure 5.4: Distributions of the relevant kinematic analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

5.3 B Jet Tagging

While the basic scheme of the tagging procedure has been kept unchanged from that previously discussed, improved tagging efficiencies were achieved for the 189 GeV data by introducing variables more sensitive to the differences between jets with b hadrons and those without. A full description of the algorithm can be found elsewhere [113] and only a brief description of the changes will be presented here.

The algorithm used to tag displaced vertices was one of those updated. This method uses an ANN to discriminate tracks originating at the event primary vertex from those originating at secondary vertices. This track-based ANN (track-ANN) was trained using Z^0-pole Monte Carlo samples. It uses as input the impact parameter of the track with respect to the primary vertex and the transverse momentum with respect to the jet axis, among other variables. The charged tracks belonging to a given jet are then ranked according to the output of the track-ANN. The most significant tracks in this ranking, up to a maximum of six, are then used to construct a vertex candidate for each jet. For each remaining track, if the χ^2 in the vertex-refitting procedure indicates the track is consistent with coming from the displaced vertex, it is added to the vertex.

In order to gain b-tagging efficiency in events where the secondary vertices are less distinct, a method to combine track impact parameters was also introduced. In this method, the impact parameter significances $S^{r\phi}$ and S^{rz}, in the $r\phi$ and rz projections, respectively, are formed by dividing the track impact parameters by their estimated errors. The distributions of $S^{r\phi}$ and S^{rz} for each quark flavour from Z^0-pole Monte Carlo are used as the probability density functions (PDF's) $P(S^{r\phi}|q)$ and $P(S^{rz}|q)$ (q=uds,c and b). The combined conditional probability for each track is then $P(S^{r\phi}, S^{rz}|q) = P(S^{r\phi}|q) \cdot P(S^{rz}|q)$. As each track is presumed to be independent, we can then define the combined estimator for each quark flavour q, $F_q = \prod_i P(S^{r\phi}, S^{rz}|q)_i$ where the product is taken over all tracks passing the quality requirements. The final estimator for the event is then $\mathcal{L}_{\rm IP} = F_b/(F_{uds} + F_c + F_b)$.

The lifetime tag is again based on an ANN. However, the five variables used in our 183 GeV analysis were replaced by the following four:

- The vertex significance likelihood ($\mathcal{L}_{\rm V}$): The likelihood for the vertex significance is computed analogously to the $\mathcal{L}_{\rm IP}$ above, using the decay length significance of the secondary vertex rather than the impact parameter significance of the tracks.
- The reduced significance likelihood ($\mathcal{R}_{\rm V}$): The track with the largest impact parameter significance with respect to the primary vertex has been removed from the secondary vertex candidate, and the remaining tracks are used to recompute the likelihood $\mathcal{L}_{\rm V}$. If the original vertex has only two tracks, the function is calculated from the impact parameter significance of the remaining track. This input variable reduces the sensitivity to mismeasured tracks.
- The combined impact parameter likelihood ($\mathcal{L}_{\rm IP}$) described above.
- The reduced impact parameter likelihood ($\mathcal{R}_{\rm IP}$): The track having the biggest impact parameter significance has been removed in the calculation of $\mathcal{L}_{\rm IP}$.

The boosted sphericity of the jet, used in previous b tag as an input to the final likelihood, has been replaced with three variables, which are combined with a separate ANN. These three inputs are: the number of energy-flow objects around the central part of the jet; the angle between the jet axis and its boosted sphericity axis; and the C-parameter for the jet boosted back to its rest frame. The high-p_t lepton tag has not been changed.

The outputs from the lifetime ANN, the jet-kinematics ANN and the high-p_t lepton tag are combined in the same manner described by Equation 4.10 to define a final B-discriminate, $\mathcal{B}$, for each jet.

The distribution of $\mathcal{B}$ can be seen in Figure 5.5 for calibration data taken at $\sqrt{s} = m_{Z^0}$ in 1998. Figure 5.5 also shows the performance of the new b-tagging compared with the previous

version [104]. The tagging efficiency for b-flavoured jets have been verified to an accuracy of 1% with a double-tagging technique using the Z^0 calibration data; a comparison of the results of this double-tagging test between the data and Monte Carlo is also shown in Figure 5.5.

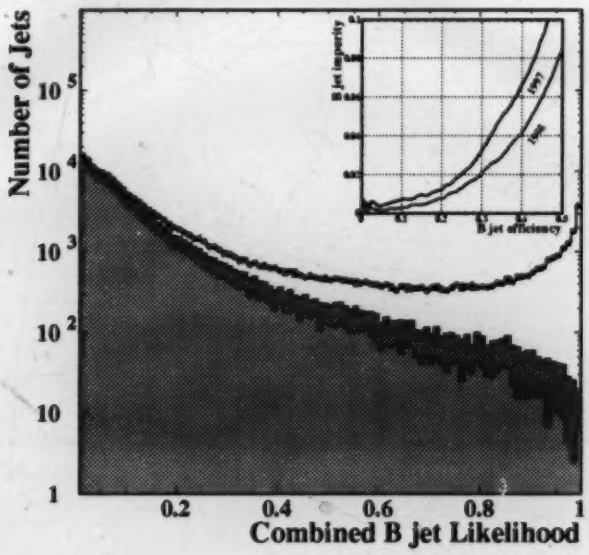


Figure 5.5: B-tagging performance and modelling. The distribution of the output of the b-tagging algorithm, $\mathcal{B}$, for jets in calibration data taken at $\sqrt{s} = m_{Z^0}$ in 1998, compared to the Monte Carlo expectation. The data distribution is given by the points, with error bars smaller than the plot symbols. The open histogram shows the distribution of $\mathcal{B}$ for b-flavour jets, and the dark (light) grey histogram shows the contribution from c (uds) flavour jets, expected in a Monte Carlo simulation. Inset: The b-tagging (1998) performance for hadronic jets in Z^0 decay compared with the previous version (1997) [104].

5.3.1 Application to 6 Jet Analysis

From experience gained in the 183 GeV analysis, it was found that the discriminating efficiency of Equation 4.10 did not strongly depend on the choice of the weight factors, α, β. With no *a priori* choice for their values, the case $\alpha = \beta = 1$, corresponding to $w_{uds} = w_c = w_b = 1$, was used. This selection follows from the *uniform prior* assumption in Bayesian theory, from which Equation 4.10 was derived, and is desirable in that the b-discriminate of the jet is then based on its observed parameters without weighting by prior expectation. The b-discriminate of the i^{th} jet is defined as,

$$\mathcal{B}_i = \frac{1}{1 + C_i + U_i}, \tag{5.1}$$

and the b-discriminate of the event is defined as,

$$\mathcal{B}_{\mathrm{Evt}} = \frac{1}{1 + \prod_i C_i + \prod_i U_i}, \tag{5.2}$$

where the products are take over the reconstructed jets. The distributions of Equations 5.1 and 5.2 for events reconstructed as six jets are shown in Figures 5.6 and 5.7, and in Figures 5.8 and 5.9 for events reconstructed as three jets.

With the improvements in the b-tagging algorithms, it was found that there was no observable benefit by including the additional vertex variables, `TDSUM,TDRSUM` (Section 4.4.3) in the analysis.

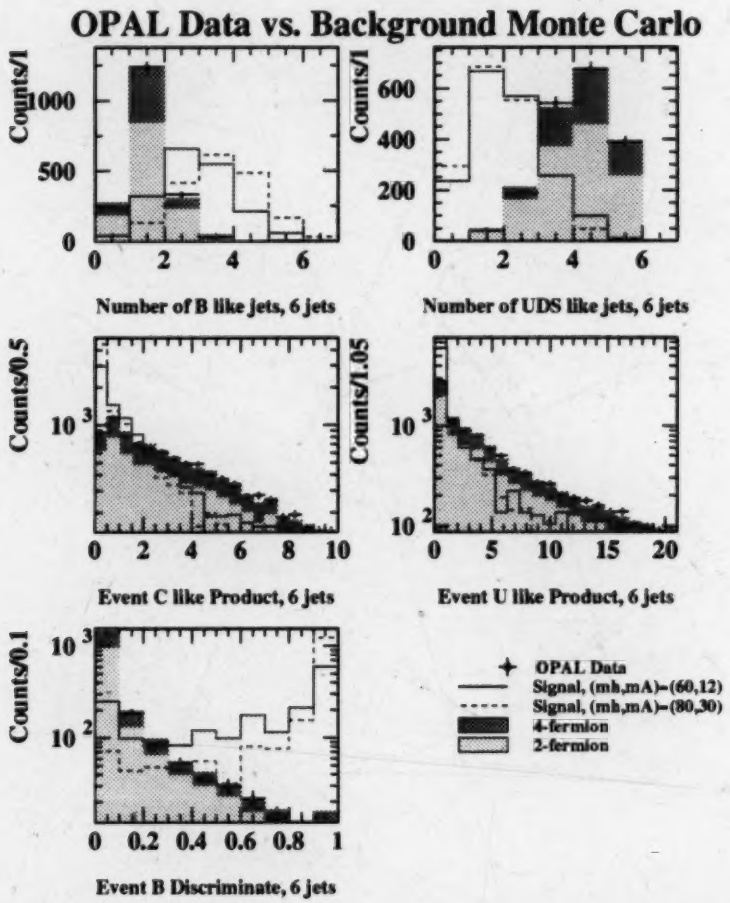


Figure 5.6: Distributions of the relevant b-tagging analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

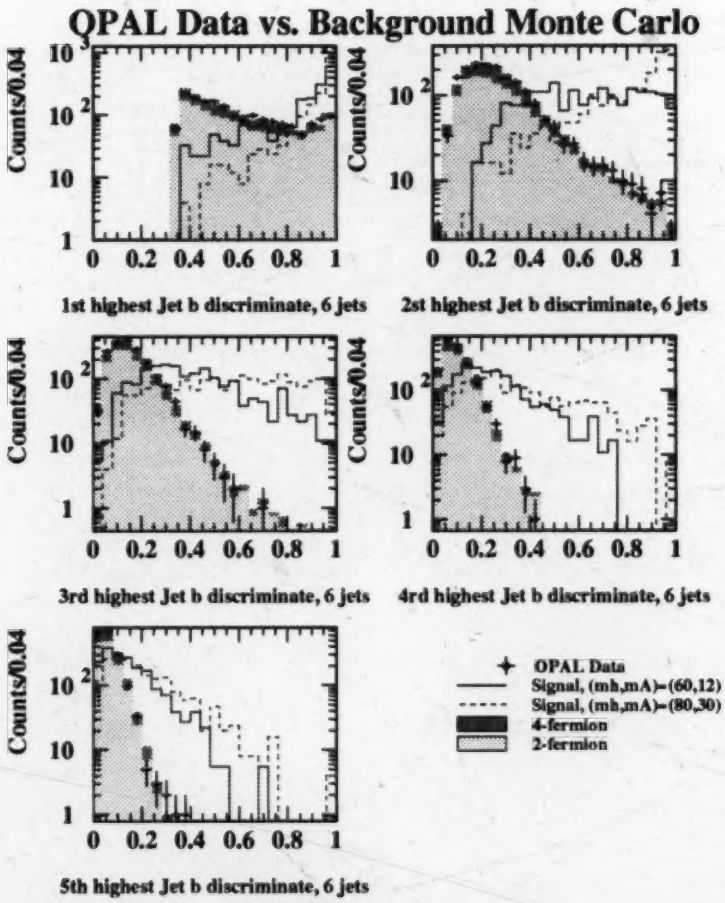

OPAL Data vs. Background Monte Carlo
Counts/0.04
1st highest Jet b discriminate, 6 jets
2st highest Jet b discriminate, 6 jets
3rd highest Jet b discriminate, 6 jets
4rd highest Jet b discriminate, 6 jets
5th highest Jet b discriminate, 6 jets
OPAL Data
Signal, (mh,mA)=(60,12)
Signal, (mh,mA)=(80,30)
4-fermion
2-fermion

Figure 5.7: Distributions of the relevant b-tagging analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

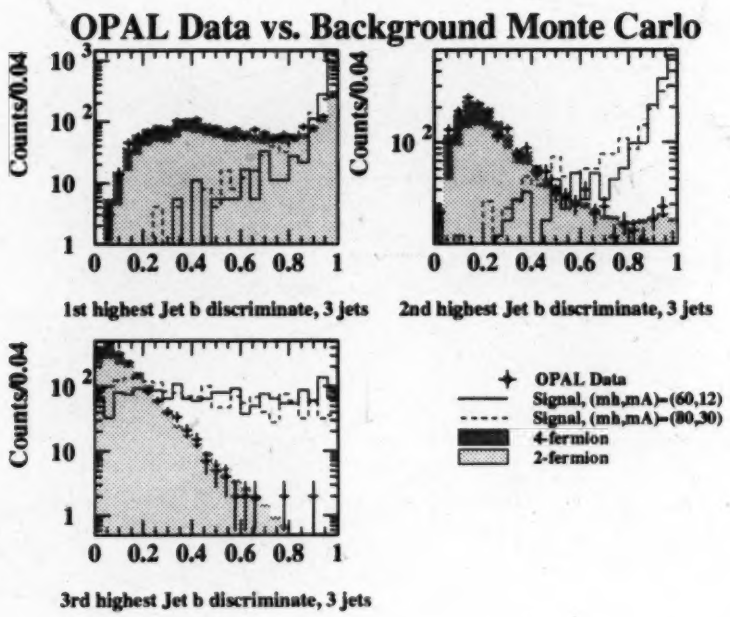


Figure 5.8: Distributions of the relevant b-tagging analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

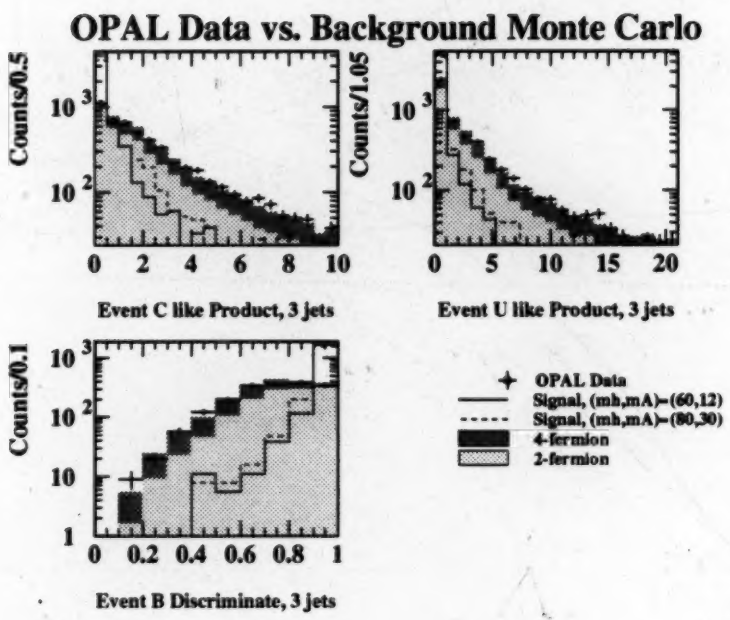


Figure 5.9: Distributions of the relevant b-tagging analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms). Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

Analysis Signal Results											
	Data	Signal Mass points									
		(24,12)	(40,12)	(60,12)	(80,12)	(40,20)	(50,20)	(70,20)	(60,30)	(80,30)	(80,40)
Total	18701	459	500	500	500	500	500	500	500	500	451
Presel	3121	418	452	451	463	453	459	490	486	490	440
Trk Mult	2298	362	341	377	446	339	331	478	452	481	437
Jet Trk Mult	1841	317	280	334	392	318	293	452	436	459	425
Jet Num.	259	182	199	258	310	237	230	364	375	413	393
$\mathcal{B}_{\mathrm{Evt}}$	7	39	56	83	149	75	100	212	208	263	256
ϵ		.085	.112	.166	.297	.150	.199	.424	.416	.526	.568
$\pm$ (Stat)		.013	.014	.017	.020	.016	.018	.022	.022	.022	.023

Table 5.2: The signal efficiencies for the Cut based 189 GeV analysis. The number of signal events surviving each cut for the 189 GeV data. Efficiencies for the signal events, denoted by the columns $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0})$, are given in the last row along with the Statistical error estimates.

5.4 Cut Based Analysis

For the 189 GeV data, a simple cut based analysis was performed in order to gauge the expected background levels for the neural network analysis, which will be developed in a later section.

The stated aim of achieving high signal efficiency over the whole of the relevant $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0})$ region makes the successful application of topological cuts difficult, so they were avoided in preference for variables based on b-tagging which exploits the presence of the many b flavoured jets expected in a signal event.

After the initial event quality and shape cuts, hard cuts were made on the b-quark content of the event:

- Given $\mathcal{C}_i(\mathcal{U}_i)$ is the ratio of probability of the i^{th} jet being c(uds)-flavoured to its probability of being b-flavoured, then $\mathcal{B}_i$, given by Equation 5.1, is the probability of the i^{th} jet being b-flavoured. Each jet was classified based on $\max\{\mathcal{B}_i, \mathcal{C}_i \cdot \mathcal{B}_i, \mathcal{U}_i \cdot \mathcal{B}_i\}$ and the event was required to have at least 2 jets which were most likely b-flavoured and less than four jets which were most likely uds-flavoured.

- The event b-discriminate, $\mathcal{B}_{\mathrm{Evt}}$, where the products in Equation 5.2 are taken over all the reconstructed jets, was selected to be greater than 0.92 .

The numbers of events passing each requirement, compared with estimates from the background simulations normalized to the integrated luminosity of $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\ \mathrm{pb}^{-1}$, are shown in Tables 5.2 and 5.3. Also shown are the detection efficiencies for various $\mathrm{e^+e^-} \to \mathrm{h^0\ A^0} \to \mathrm{A^0\ A^0\ A^0} \to 3(\mathrm{b\bar{b}})$ signal mass points.

7 events pass the selection requirements, consistent with the background expectation of $8.08 \pm^{\mathrm{Stat.}}$ 0.88 events. Table 5.4 lists the Run end Event numbers of those events in the 189 GeV OPAL data which passed this analysis. Systematic errors where not calculated for this analysis as it was only used to check the results of the Neural Network analysis, described below.

Analysis Background Results						
	Data	$(Z/\gamma)^* \to q\bar{q}$	$e^+e^- \to q\bar{q}q\bar{q}$	$e^+e^- \to q\ell q\ell$	$e^+e^- \to q\ell\nu\ell$	$\langle n \rangle$
Total	18701	16976.4 (148604)	1447.9 (27088)	1492.8 (43396)	4388.8 (99500)	24306.0
Presel	3121	2155.5 (18868)	556.69 (10415)	124.36(3615)	53.06 (1203)	2889.58
Trk Mult	2298	1502.45(13152)	509.57(9533)	83.86(2438)	31.75 (720)	2127.64
Jet Trk Mult	1841	1204.96(10548)	466.46 (8727)	46.75(1359)	21.75(493)	1739.91
Number of Jets	259	76.84(1548)	53.02 (992)	2.64 (77)	3.61(82)	236.12
$\mathcal{B}_{Evt}$	7	5.61 ± 0.80(49)	2.16 ± 0.34(40)	0.05 ± 0.04(1)	0.26 ± 0.11(6)	8.08 ± 0.88

Table 5.3: The predicted Standard Model background for the cut based 189 GeV analysis. The number of predicted background Monte Carlo events surviving each cut for the 189 GeV data and the expected background, $\langle N \rangle$, normalized to $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50$ pb^{-1}. Bracketed values are the number of events passing the cuts from the Monte Calro data sample. Errors are Statistical only.

Data Events Selected

Run	Event
9481	57123
9571	29348
9579	49100
9824	22427
10209	56167
10248	24704
10259	43333

Table 5.4: Event numbers of those data events selected by the Cut based 189 GeV analysis.

5.5 Neural Network Based Analysis

As demonstrated in Figures 5.1 to 5.9, the topology of the signal events change dramatically between the allowed mass regions. However, many topological variables are seen to still contain discriminatory information which cannot be efficiently used in the a standard cut based analysis. This prompted a more detailed search for $e^+e^- \to h^0 A^0 \to A^0 A^0 A^0 \to 3(b\bar{b})$ to be performed. In this analysis, candidates are selected using two neural networks: one combines kinematic and topological variables; the other exploits heavy flavour tagging.

The previous 183 GeV ANN analysis was optimized at specific signal mass points and performed well at the training mass but were not efficient for all signal masses. For this analysis, the neural networks were trained with signal events from across the relevant (m_{h^0}, m_{A^0}) range. The effect being that the acceptance at a single mass point was lower than that of a dedicated neural network, but the efficiency of the single network was now acceptable for all signal masses.

The networks were constructed and trained as described in Section 4.5.1. The specific network topologies of were course different, which will be discussed below, and the frequency of updating during the training phase was changed. In the previous analysis the weights were updated with every 10 training patterns processed. In this case, the training "signal" is in fact a randomly ordered mixture of signal events generated at eight different mass points. To prevent the summed error measure, used in the updating algorithm, from being based on only a small number of the mass points, the frequency of updating was increased to every 25 training patterns processed.

A number of other network topologies were also studied including: a monolithic network as used in the previous 183 GeV analysis; networks with more than a single hidden layer; networks with more than one output node; and training identical networks at each mass point and subsequently using their outputs as inputs to a final neural network. None of these schemes provided any significant efficiency improvement over the others and, in the end, the analysis presented here was chosen from observing trends in the networks studied with the goal of using the most simple system possible.

It was found that networks which included both kinematic and heavy flavour information invariably trained to emphasize the heavy flavour component, which was the most consistent feature between all the signal points. By training a network on kinematic variables specifically, the network was forced to look deeper for discriminatory information contained in them. A separate network emphasizing heavy flavour tagging was then trained, and both networks applied to maximize the background rejection while keeping the signal efficiencies as large as possible.

To begin, events were first required to pass the preselection cuts described in Section 5.2.4.

5.5.1 Training and Testing data sets

The available Monte Carlo was split into two subsets: the training data, which was used to train and validate the networks; and, the analysis data, used to determine the efficiencies quoted as the results of this analysis. The training subset, listed in Table 5.5, was further subdivided into 80% for training phase and 20% for the verification phase.

With one exception, the 2.5K events in each signal data sample listed in Table 5.1, were used for training and verification. To maximize the data available for training, the 20%, or 500 events used for the verification phase in each signal set, were reused in the evaluation phase of the analysis. The mass point $(m_{h^0}, m_{A^0}) = (80, 40)$ was not used in training the networks as it is efficiently identified by networks trained to identify $(m_{h^0}, m_{A^0}) = (80, 30)$.

A subset of the `PYTHIA, GRC4F` [1] background sample listed in Table 5.1 was used for training and verification purposes. The remaining data formed the Monte Carlo sample used to evaluate the expected background in this analysis. The training sample was mixed based on the relative effective cross-sections (cross-sections multiplied by preselection efficiency) of the contributing channels. In

[1] $e^+e^- \to q\bar{q}e^+e^-$ data was not used in training as its effective cross-section was too low to contribute to the final expected background.

this case, only the channel with the highest effective cross-section would use 80% of its available data for the training phase. The remaining channels contribute, based on their effective cross-section relative to the highest one, with all the remaining data being used in the verification set.

Additional Preselections

To prevent the networks from learning characteristics of those events easily rejected by simple cuts, the training data samples were filtered through a simple preselection. In addition to those cuts described in Section 5.2.4, the training events were additionally required to pass:

- The most significant jet, in decreasing order of the value of its b-discriminate, $\mathcal{B}_{\text{Jet}}$, described in Equation 5.1, was required to have $\mathcal{B}_{\text{Jet}} \geq 0.5$.
- $\log(\prod_{i=1}^{6} \mathcal{C}_i) \leq 7.0$ and $\log(\prod_{i=1}^{6} \mathcal{U}_i) \leq 7.0$ where $\mathcal{C}_i, \mathcal{U}_i$, are defined in Equation 5.2.

These additional cuts were not applied to the data in the main analysis and, in fact, were found to have no impact on the final results, supporting the argument that once the network had been sufficiently trained to identify signal-like events, the presence of additional background with low b content was easily rejected.

Process	File	Number of Events	σ (pb)
HZHA generated signal, $Zh \rightarrow 3$(bb)			
$m_{h^0} = 40$, $m_{A^0} = 12$	...L20012/P189HA/R8444	2500(1400)	
$m_{h^0} = 40$, $m_{A^0} = 20$	...L20012/P189HA/R8445	2500(1390)	
$m_{h^0} = 50$, $m_{A^0} = 20$	...L20012/P189HA/R8446	2500(1397)	
$m_{h^0} = 60$, $m_{A^0} = 12$	...L20012/P189HA/R8447	2500(1534)	
$m_{h^0} = 60$, $m_{A^0} = 30$	...L20012/P189HA/R8448	2500(2074)	
$m_{h^0} = 70$, $m_{A^0} = 20$	...L20012/P189HA/R8449	2500(2103)	
$m_{h^0} = 80$, $m_{A^0} = 12$	...L20012/P189HA/R8450	2500(1957)	
$m_{h^0} = 80$, $m_{A^0} = 30$	...L20012/P189HA/R8451	2500(2232)	
Main PYTHIA, GRC4F background			
$(Z/\gamma)^* \rightarrow q\bar{q}$	...L20012/P189QQ/R5078	155095(4609)	2.934
$e^+e^- \rightarrow q\bar{q}q\bar{q}$	...L20012/GRC4F/R7846	15000(910)	0.511
$e^+e^- \rightarrow q\ell q\ell$	...L20012/GRC4F/R8055	43396(173)	0.035

Table 5.5: Monte Carlo signal and PYTHIA, GRC4F background data sample used in the training of the ANNs. The values in the Events column indicate the total sample size considered before preselection, those in parentheses indicate the number of events passing the preselections and thus are the total number used in the training of the ANN. The cross-sections in the last column are the effective cross-sections used to mix the background sets during training.

5.5.2 Kinematic Neural Network

Recall the relevant (m_{h^0}, m_{A^0}) range can be split into two regions: $m_{h^0} + m_{A^0} > m_{Z^0}$ and $m_{h^0} + m_{A^0} < m_{Z^0}$. In the latter region, radiative return events are quite common and no cuts on $\sqrt{s'}$, C or D parameters, or a number of other topological variables were possible. However, many of these topological variables were seen to still contain discriminatory information, on a mass point by mass point analysis, which was not efficiently used in the cut based analysis described above. It was felt that a neural network analysis could exploit the discriminatory information in these variables as the method is well suited for untangling complicated correlations without using a hard selection criteria.

The first network trained with 14 input nodes, 1 hidden layer with 50 hidden nodes and 1 output node, using the following variables as inputs:

- Standard event quantities: thrust, sphericity, oblateness and the C and D parameters.
- The $2^{nd} - 6^{th}$ normalized Fox-Wolfram [100] moments, and
- the logarithms of y_{32}, y_{43}, y_{54} and y_{65} where where $y_{n+1,n}$ is the jet resolution parameter at which events change from $n+1$ to n jets in the Durham scheme.

The distributions of these inputs, after the preselection cuts in Section 5.2.4 were applied, are shown in Figures 5.1 to 5.4

Network Pruning

Initial training used a network with fifty units in the hidden layer. It is expected that the network can be pruned of hidden units without significantly effecting the generalization error of the final system. To test this without needing to retrain a number of different networks, the Principal Component Pruning (PCP) analysis described in Chapter 3 was applied.

Table 5.6 presents the twenty most significant eigenvalues from the Principal Component Analysis (PCA) applied to the outputs of the fifty units in the hidden layer. Essentially, all the discriminatory information is contained in these first twenty eigenvalues, implying the original choice of fifty hidden units was indeed more complex than needed.

Principal Component Pruning is applied by successively projecting out the contributions of the least significant eigenvectors in the PCA space and then rotating back to the space of the hidden layer outputs to reset the internal weights and thresholds connecting the hidden layer to the output nodes. The network, with its PCP adjusted weights, is then re-run on the validation data set to calculate the new generalized error. This error is compared to the its original value, and a new network topology is chosen based on the number of eigenvectors in the PCA space which are seen not to contribute significantly to the performance of the network.

Table 5.7 presents the results on the current network topology. The left hand column indicates the number of ordered principal component eigenvectors remaining after the projection was applied. The right hand column shows the effect on the projection on the generalized error, calculated by applying the network to the validation data set. We see that projecting out the thirty least significant eigenvectors has little effect on the generalized error of the network, indicating that a network trained with fewer than fifty hidden units should be sufficient to efficiently classify the data classes.

Direct confirmation of the PCP analysis was then required to make the final selection of the pruned network topology. Using the results in Table 5.7 as a guide, networks were trained with 5, 10, 20 and 30 units in the hidden layer. Figure 5.10 directly compares the $(m_{h^0}, m_{A^0}) = (80, 30)$ GeV efficiency of these networks for a given expected background level. From this we see that the original network with fifty hidden units was significantly different than those with fewer nodes, indicating the network may have been too complex. In this case, the network was chosen to have thirty hidden nodes, which is large enough to maintain good signal efficiency while not being needlessly complex.

Eigenvalue	λ	% of Trace	cum. % of Trace
1	2.1904	47.1772	47.1772
2	1.3724	29.5589	76.7360
3	0.3915	8.4312	85.1673
4	0.3098	6.6726	91.8399
5	0.1424	3.0666	94.9065
6	0.0750	1.6149	96.5214
7	0.0531	1.1431	97.6646
8	0.0321	0.6920	98.3566
9	0.0237	0.5094	98.8660
10	0.0126	0.2706	99.1366
11	0.0102	0.2191	99.3557
12	0.0078	0.1680	99.5237
13	0.0059	0.1273	99.6510
14	0.0055	0.1195	99.7705
15	0.0037	0.0791	99.8496
16	0.0031	0.0670	99.9166
17	0.0013	0.0272	99.9437
18	0.0008	0.0166	99.9604
19	0.0007	0.0150	99.9753
20	0.0004	0.0085	99.9839

Table 5.6: The twenty most significant eigenvalues from the Principal Component Analysis applied to the outputs of the fifty units in the hidden layer. Essentially all the discriminatory information is contained in these first twenty eigenvalues, implying the original choice of fifty hidden units was more complex than needed.

Network Selection

The final network was trained with 14 input nodes, 1 hidden layer with 30 hidden nodes and 1 output node. Figure 5.11 presents the training evolution of this network. Plot (a) shows the generalization error evolution for the signal and background channels. After approximately 8000 epochs, the errors have levelled off and we see, in plot (b), that the hidden layer saturation, a measure of the learning rate, has slowed. Plot (c) shows the efficiency evolution, here for $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV with an arbitrary cut of 0.8 on the network output, has reached a level plateau with little evidence for over training after 8000 epochs. The final plot, (d), presents normalized plots of the network output for a sample signal $(m_{h^0}, m_{A^0}) = (60, 30)$ GeV and the large background channel, $(Z/\gamma)^* \to f\bar{f}$, after 8000 epochs. From these observations, the network state after 8000 epochs was used as the network in the final analysis.

Events were required to have a kinematic network output greater than 0.68 . The distribution of the network output, after preselections have been applied, can be seen in Figure 5.15a.

Eigenvectors Used	% Change in Generalised Error
50	0.000
20	0.001
15	0.006
10	0.335
9	0.354
8	1.596
7	4.180
6	6.858
5	7.450
4	35.672
3	39.139
2	40.136
1	40.075

Table 5.7: Principal Component Pruning applied to the outputs of the units in the hidden layer of the Kinematic ANN. The left hand column indicates the number of ordered principal component eigenvectors remaining after the projection was applied. The right hand column shows the effect of the projection on the generalized error, calculated by applying the network to a data set not used in the training phase. We see that projecting out the thirty least significant eigenvectors has little effect on the generalized error of the network, indicating that a network trained with fewer than fifty hidden units should be sufficient to efficiently classify the data classes.

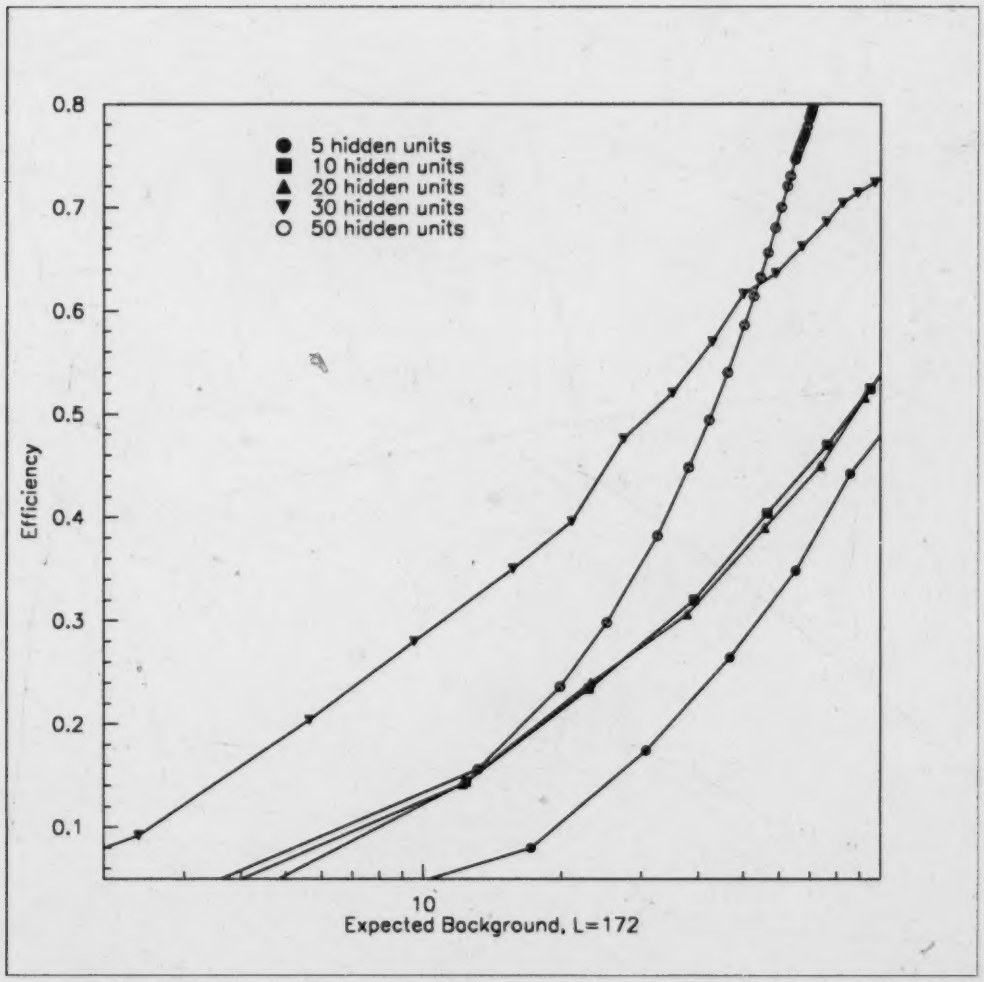


Figure 5.10: Direct comparision of Kinematic Neural Networks trained for Principal Component Pruning analysis.

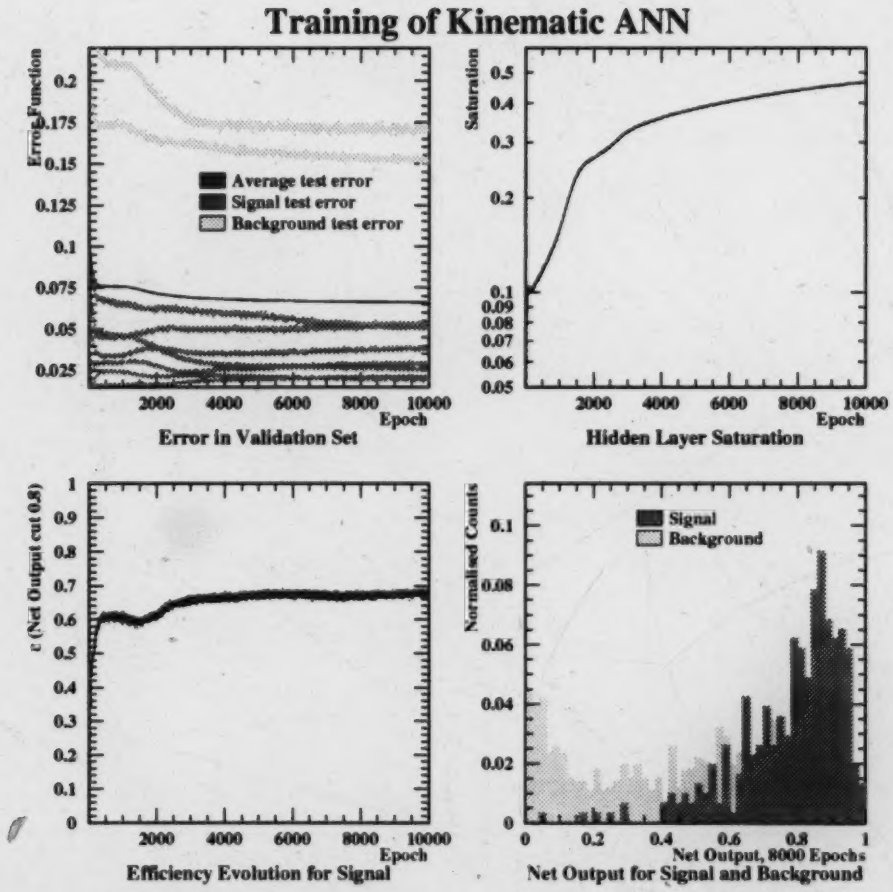


Figure 5.11: Evolution of the Kinematic Neural Network training. Plot (a) shows the generalization error evolution for the signal and background channels. After approximately 8000 epochs, the errors have levelled off and we see, in plot (b), that the hidden layer saturation, a measure of the learning rate, has slowed. Plot (c) shows the efficiency evolution, here for $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 30)$ GeV, for an arbitrary cut of 0.8 on the network output has reached a level plateau with little evidence for over training after 8000 epochs. The final plot, (d), presents normalized plots of the network output for a sample signal $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 30)$ GeV and the large background channel, $(\mathrm{Z}/\gamma)^* \to \mathrm{f\bar{f}}$, after 8000 epochs.

5.5.3 B-Tagging Neural Network

Once the kinematic information was exhausted, a final B-tagging network was applied. As mentioned above, the topology of the signal events change dramatically between the regions $m_{h^0} + m_{A^0} > m_{Z^0}$ and $m_{h^0} + m_{A^0} < m_{Z^0}$. Depending on the (m_{h^0} , m_{A^0}) mass point and whether radiative return occurred, many of the signal events could be better described as three jets, where the $A^0 \to b\bar{b}$ decays are sufficiently boosted so as to look like a single jet. At higher (m_{h^0} , m_{A^0}) mass points the three $A^0 \to b\bar{b}$ decays can be well separated and identified. Relying, then, on the ability of the network to untangle correlated inputs and noting the above observation, a second network was trained with 15 input nodes, 1 hidden layer with 50 hidden nodes and 1 output node, using the following three and six jet B-tagging variables as inputs:

- $\mathcal{B}_i$ for three jets and the largest four $\mathcal{B}_i$ for six jets.
- $\mathcal{B}_{Evt}$, as well as the logarithms of $\prod_i \mathcal{C}_i$ and $\prod_i \mathcal{U}_i$, for both the three and six jets cases
- Additionally, $\sqrt{s'}$ was included to help the network separate the three and six jet cases.

Network Pruning

The hidden layer was pruned in the same fashion as described above. Tables 5.8 and 5.9 present the results of the PCP analysis applied to the outputs of the fifty units in the hidden layer of the $\mathcal{B}$-tagging network. It is shown that a network with as few as twenty units in the hidden layer should contain almost all the discriminatory information.

Eigenvalue	λ	% of Trace	cum. % of Trace
1	1.3809	75.6270	75.6270
2	0.1472	8.0607	83.6877
3	0.0995	5.4493	89.1370
4	0.0633	3.4654	92.6024
5	0.0491	2.6875	95.2899
6	0.0265	1.4510	96.7410
7	0.0196	1.0715	97.8125
8	0.0153	0.8366	98.6491
9	0.0092	0.5057	99.1548
0	0.0057	0.3121	99.4669
11	0.0039	0.2113	99.6782
12	0.0029	0.1563	99.8345
13	0.0014	0.0787	99.9131
14	0.0010	0.0560	99.9692
15	0.0002	0.0125	99.9817
16	0.0001	0.0051	99.9868
17	0.0000	0.0023	99.9891
18	0.0000	0.0018	99.9909
19	0.0000	0.0013	99.9922
20	0.0000	0.0010	99.9932

Table 5.8: The twenty most significant eigenvalues from the Principal Component Analysis applied to the outputs of the 50 units in the hidden layer of the $\mathcal{B}$-tagging network. Essentially all the discriminatory information is contained in these first twenty eigenvalues, implying the original choice of 50 hidden units was more complex than needed.

Eigenvectors Used	% Change in Generalised Error
50	0.000
20	0.000
15	0.000
10	0.005
9	-0.007
8	0.066
7	0.062
6	0.045
5	0.045
4	0.061
3	0.115
2	0.196
1	0.395

Table 5.9: Principal Component Pruning applied to the outputs of the units in the hidden layer of the b-tagging ANN. The left hand column indicates the number of ordered principal component eigenvectors remaining after the projection was applied. The right hand column shows the effect of the projection on the generalized error, calculated by applying the network to a data set not used in the training phase. We see that projecting out the forty least significant eigenvectors has little effect of the generalized error of the network, indicating that a network trained with fewer than fifty hidden units should be sufficient to efficiently classify the data classes.

This prediction was directly confirmed in Figure 5.12 which compares the (m_{h^0}, m_{A^0}) = (60, 30) GeV and (80, 30) GeV efficiency, of a select number of networks, for a given expected background level.

Network Selection

The final network trained with 15 input nodes, 1 hidden layer with 20 hidden nodes and 1 output node. Figure 5.13 presents the training evolution of this network and, through a similar method as discussed in Section 5.5.2, the network state after 5000 epochs was used as the network in the final analysis.

Events were required to have a b-tagging network output greater than 0.92 . The distribution of the network output, after preselections in Section 5.2.4 have been applied, can be seen in Figure 5.15b.

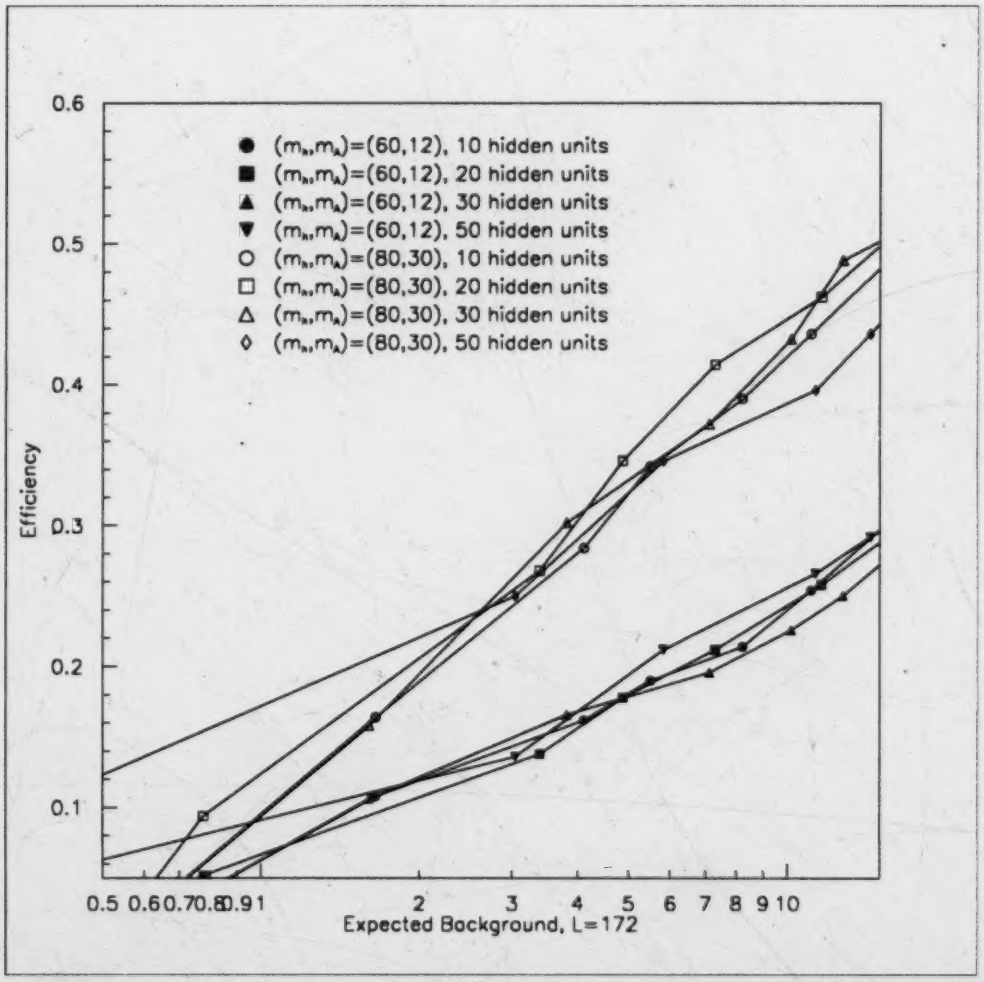


Figure 5.12: Direct comparision of b-tagging Neural Networks trained for Principal Component Pruning analysis. It is clear that networks trained with significantly fewer than fifty units in the hidden layer have signal efficiencies similar to the largest networks. In this case, the network was chosen to have 20 hidden nodes, which is large enough to maintain good signal efficiency while not being needlessly complex.

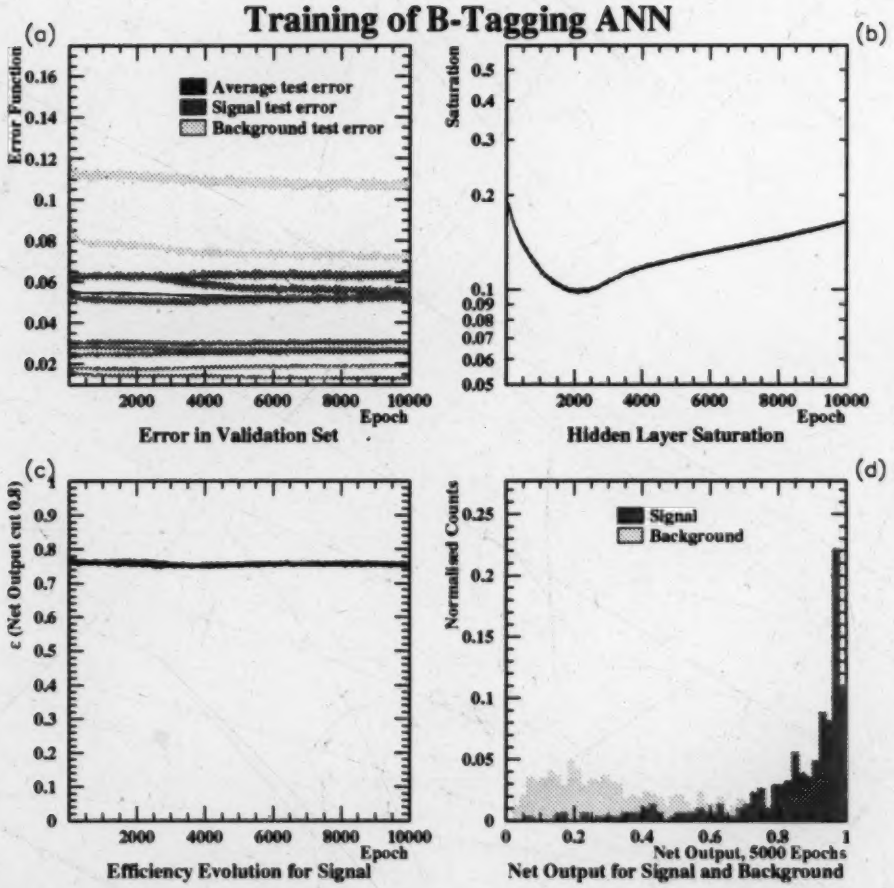


Figure 5.13: Evolution of the B-tagging Neural Network training. Plot (a) shows the generalization error evolution for the signal and background channels. After approximately 5000 epochs, the errors have levelled off and we see, in plot (b), that the hidden layer saturation, a measure of the learning rate, has slowed. Plot (c) shows the efficiency evolution, here for $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 30)$ GeV, for an arbitrary cut of 0.8 on the network output has reached a level plateau with little evidence for over training after 5000 epochs. The final plot, (d), presents normalized plots of the network output for a sample signal $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0}) = (60, 30)$ GeV and the large background channel, $(\mathrm{Z}/\gamma)^* \to \mathrm{f\bar{f}}$, after 5000 epochs.

5.5.4 Input Pruning

The obvious questions about this analysis are, " Are all the all input variables really needed? Could some be pruned?" In this case, two networks were used, each of moderate size. The seemingly large number of inputs in both networks were to address the issue of the large variance in event topology over the mass range considered ($m_{h^0} \geq 2m_{A^0}$, $m_{h^0} + m_{A^0} > m_{Z^0}$ and $< m_{Z^0}$), so it was desirable to give the network a large parameter space to generalize efficiently for the high and low mass cases.

A moderate amount of pruning could likely have been accomplished. However, the networks found their minimums quickly, $\lesssim 8000$ epochs, and were trained long enough, 10000 epochs, to have identified the onset of over-training, indicating that the selected networks were not overly complex and could be trained with the available data sets.

5.5.5 Systematic Checks

Preselection

Varying the preselection cuts by an amount equal to the difference between the means of their distributions in the data and Monte Carlo had no effect on the signal efficiencies of expected background so no systematic error was assigned.

Si Resolution Smearing

To estimate the $r\phi$ and z track resolution uncertainties due to imperfection in the Monte Carlo modelling, parametrized track smearing was applied with the OPAL routine `ODSMGL` with smearing parameters between 0.95-1.05 in $r\phi$ and a slightly larger range, 0.9-1.1, in z. These smearing parameters are scale factors so that a smearing of 1.0 would have no effect, while a value of 0.95 increases and 1.05 degrades the resolution uncertainties by 5%. With $(r\phi, z)$ smearing parameters of $(0.95, 0.9)$ and $(1.05, 1.1)$, the background prediction was 7.43 ± 1.00 and 10.72 ± 1.11, respectively, where the errors shown are statistical. The effect of this smearing on the signal is tabulated in Table 5.10 and on the background in Table 5.11.

Silicon hit dropping has been investigated using the `ODSIZR` parameters available through `ODSMGL`. It was found that the number of $r\phi$ was slightly overestimated in Monte Carlo and the number of z hits was slightly underestimated, thus the parametrized hit dropping was varied using the $r\phi$ hit-dropping scale parameter and z hit-dropping scale parameter. To bring the data in better agreement with the data, the hit dropping scale parameters applied were were 1.05 for $r\phi$ and 0.99 for z. The results were consistent with past experience were it has been shown that the sensitivity to hit dropping is much less than that due to track resolution smearing.

The relative systematic error, due to track resolution uncertainties, amounted to $\approx 15\%$ for the signal efficiency and 23.7% for the expected background prediction.

ANN Analysis: Effect of Track Smearing on Signal Results									
	Signal Mass points								
	(24,12)	(40,12)	(60,12)	(80,12)	(40,20)	(50,20)	(70,20)	(60,30)	(80,30)
1. Track resolution improved by 5% in $r\phi$, 10% in z									
3. B-Tagging Net	48	84	132	216	85	116	225	233	252
ϵ	.104	.167	.265	.433	.170	.232	.449	.466	.503
$\pm$ (Stat)	.014	.017	.020	.022	.017	.019	.022	.022	.022
1. Track resolution degraded by 5% in $r\phi$, 10% in z									
3. B-Tagging Net	40	80	116	208	75	108	220	224	246
ϵ	.086	.160	.232	.416	.151	.216	.440	.448	.492
$\pm$ (Stat)	.013	.016	.019	.022	.016	.018	.022	.022	.022

Table 5.10: Effect of Track Smearing on the signal efficiencies for the neural network based 189 GeV analysis. The number of signal events surviving each cut for the 189 GeV data. Efficiencies for the signal events, denoted by the columns (m_{h^0} , m_{A^0}), are given in the last row along with the statistical and systematic error estimates.

ANN Analysis: Effect of Track Smearing on Background Results			
	$(Z/\gamma)^* \to q\bar{q}$	$e^+e^- \to q\bar{q}q\bar{q}$	$\langle n \rangle$
1. Track resolution improved by 5% in $r\phi$, 10% in z			
3. B-Tagging Net	6.73 ± 0.98 (47)	0.70 ± 0.19 (13)	7.43 ± 1.00
1. Track resolution degraded by 5% in $r\phi$, 10% in z			
3. B-Tagging Net	9.41 ± 1.09 (75)	1.31 ± 0.21 (38)	10.72 ± 1.11

Table 5.11: Effect of Track Smearing on the predicted Standard Model background for the Neural Network based 189 GeV analysis. OPAL data compared to the number of predicted background Monte Carlo events surviving each cut and the expected background, $\langle N \rangle$, normalized to $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\ \mathrm{pb}^{-1}$. Bracketed values are the number of events passing the cuts from the Monte Calro data sample. Errors are statistical and systematic, respectively.

Different Monte Carlo Generators

The background estimates are subject to possible mismodelling of higher-order QCD processes, including those in radiative events, and of the high-end tail of the charged track multiplicity distribution.

To estimate all these sources of systematic errors related to the modelling of the Standard Model physics, the $(Z/\gamma)^* \rightarrow q\bar{q}(\gamma)$ background calculated using the HERWIG Monte Carlo was compared with that from PYTHIA. The backgrounds from the two models are consistent within the statistical error. The background from four-fermion processes, $q\bar{q}\,q\bar{q}$, predicted by GRC4F was compared with that from PYTHIA and also found to agree to within the statistical error. The predicted background using this alternative Monte Carlo sample is 7.10 ± 1.03. No alternate Monte Carlo is available for the signal events.

The uncertainty in the b-quark fragmentation function was accounted for by assuming different energy spectra for the weakly decaying b-hadrons. The energy spectra were varied such that the mean value of the distribution was shifted by $\pm$ 400 MeV [114]. Results of this $\pm$400 MeV shift in b-hadron energy spectra on the signal and predicted background are shown in Tables 5.12 and 5.13.

Transformation of Monte Carlo distributions

As discussed previously in this thesis, a transformation of all parameters relevant to an analysis can be used to determine the total systematic error in a way which accounts for all the inherent sources of systematic error, both known and unknown.

One caveat is that if the data distribution is consistent with a signal plus background hypothesis, rather than a background only hypothesis, a background subtraction method should be used first so that the transformation of the Monte Carlo background will not include artifacts of the signal events. In this analysis we see no significant excess in the data which leads us to believe that we may proceed with the transformations as a total systematic error study.

Figure 5.14 presents the effect of this Monte Carlo transformation technique applied to a sample of the parameter distributions used in this analysis. The transformation had little effect, $\mathcal{O}(1\%)$, on the signal efficiencies, as presented in Table 5.14. This can be explained by observing that signal distributions which passed the analysis are typically out of the region of the Data vs. Monte Carlo fitting used to define the transformations, in which case the transformation is reduced to a unit mapping.

The transformation applied to the the background samples, Table 5.15, yields a predicted background of 7.16 ± 0.87 events, which is consistent with the other checks. This value was not used in the calculation of the total systematic error but is presented as a check that no extreme differences between data and Monte Carlo seem to exist.

ANN Analysis: Effect of ± 400 MeV shift in b-hadron energy spectra on Signal Results									
	Signal Mass points								
	(24,12)	(40,12)	(60,12)	(80,12)	(40,20)	(50,20)	(70,20)	(80,30)	(80,40)
$+400$ MeV shift in b-hadron energy spectra									
3. B-Tagging Net	58	90	130	226	105	129	248	277	253
ϵ	.126	.181	.260	.452	.209	.259	.495	.554	.560
$\pm$ (Stat)	.015	.017	.019	.022	.018	.019	.022	.022	.023
-400 MeV shift in b-hadron energy spectra									
3. B-Tagging Net	35	78	106	179	90	107	211	228	247
ϵ	.075	.156	.212	.357	.181	.213	.422	.457	.547
$\pm$ (Stat)	.012	.016	.018	.021	.017	.018	.022	.022	.023

Table 5.12: Effect of ± 400 MeV shift in b-hadron energy spectra on the signal efficiencies for the neural network based 189 GeV analysis. The number of signal events surviving each cut for the 189 GeV data. Efficiencies for the signal events, denoted by the columns $(m_{\mathrm{h}^0}, m_{\mathrm{A}^0})$, are given in the last row along with the statistical and systematic error estimates.

ANN Analysis: Effect of ± 400 MeV shift in b-hadron energy spectra on the Background Results			
	$(Z/\gamma)^* \to q\bar{q}$	$e^+e^- \to q\bar{q}q\bar{q}$	$\langle n \rangle$
$+400$ MeV shift in b-hadron energy spectra			
3. B-Tagging Net	8.00 ± 0.96 (70)	0.91 ± 0.22 (17)	9.04 ± 0.98
-400 MeV shift in b-hadron energy spectra			
3. B-Tagging Net	7.61 ± 0.93 (67)	0.79 ± 0.21 (15)	8.56 ± 0.96

Table 5.13: Effect of ± 400 MeV shift in b-hadron energy spectra on the predicted Standard Model background for the neural network based 189 GeV analysis. OPAL data compared to the number of predicted background Monte Carlo events surviving each cut and the expected background, $\langle N \rangle$, normalized to $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50$ pb^{-1}. Bracketed values are the number of events passing the cuts from the Monte Carlo data sample. Errors are statistical and systematic, respectively.

ANN Analysis: Effect of MC Transformation on Signal Results										
	Signal Mass points									
	(24,12)	(40,12)	(60,12)	(80,12)	(40,20)	(50,20)	(70,20)	(60,30)	(80,30)	(80,40)
3. B-Tagging Net	43	80	116	196	97	110	217	212	243	240
ϵ	.094	.160	.232	.393	.194	.221	.433	.424	.486	.532
$\pm$ (Stat)	.014	.016	.019	.022	.018	.018	.022	.022	.022	.023

Table 5.14: Effect of MC Transformation on the signal efficiencies for the neural network based 189 GeV analysis. The number of signal events surviving each cut for the 189 GeV data. Efficiencies for the signal events, denoted by the columns (m_{h^0} , m_{A^0}), are given in the last row along with the statistical and systematic error estimates.

ANN Analysis: Effect of MC Transformation on Background Results			
	$(Z/\gamma)^* \to q\bar{q}$	$e^+e^- \to q\bar{q}q\bar{q}$	$\langle n \rangle$
3. B-Tagging Net	6.31 ± 0.85 (55)	0.70 ± 0.19 (13)	7.16 ± 0.87

Table 5.15: Effect of MC Transformation on the predicted Standard Model background for the Neural Network based 189 GeV analysis. OPAL data compared to the number of predicted background Monte Carlo events surviving each cut and the expected background, $\langle N \rangle$, normalized to $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\ \mathrm{pb}^{-1}$. Bracketed values are the number of events passing the cuts from the Monte Carlo data sample. Errors are statistical and systematic, respectively.

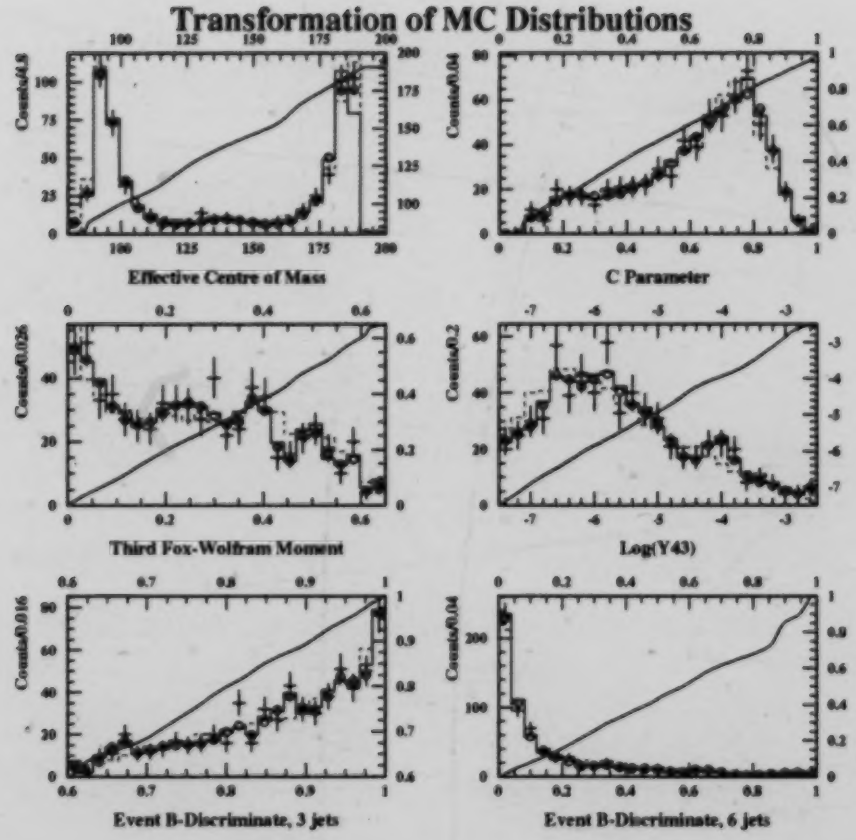


Figure 5.14: Sample of analysis variables after the Monte Carlo Transformation technique was applied. The spline fit (open circles) to the raw data distribution (crossed points) is the target for the raw Monte Carlo Background (dashed histogram). The Solid histogram is the transformed Monte Carlo distribution.
Overlayed on the histogram is a plot of the generated Monte Carlo transformation function, to be read with the upper and right axes. Had there been no difference between the data and Monte Carlo distributions, this function would be a unit mapping. Deviations from a simple $y = x$ line indicate differences not modelled by the monte carlo which will contribute to the overall systematic uncertainty calculated with this method.

Systematic Errors Applied

The overall uncertainty on the detection efficiencies for the signal are dominated by the statistics of the 500-event Monte Carlo samples, typically contributing 5% relative error. The relative systematic error on the efficiency due to the track parameter resolution uncertainty was found to be $\approx 15\%$. The systematic error due to the uncertainty in the b-hadron decay multiplicity was found to be $\approx 10\%$. The total systematic error in the signal efficiencies are typically 15%.

The systematic errors on the background estimates similarly include contributions from the modelling of the distributions used in the event selection and from Monte Carlo statistics. To account for possible mismodelling of higher-order QCD processes affecting track multiplicity, jet formation and the hadronization phenomena, the predicted background was checked using alternative Monte Carlo sample leading to a predicted background of 7.10 ± 1.03. The systematic errors due to the b-hadron multiplicity modelling and track resolutions were estimated to be 5.6% and 23.7%, respectively, using the same methods as were used to calculate the systematic errors on the detection efficiencies. Combined with the statistical error of the Monte Carlo prediction (11%), the overall systematic error on the background amounts to 29%. The Monte Carlo transformation method, which takes into account all errors contributing to the disagreement between data and Monte Carlo, predicts a systematic error on the background of $\approx 20\%$.

5.6 Results

Summarizing then, the analysis applied was:

1.
   - The LEP Multihadronic event selection flag was set
   - The polar angle of the thrust axis, θ_T, is required to satisfy $|\cos\theta_T| \leq 0.9$.
   - Charged track multiplicity ≥ 20.
   - The C-parameter is required to be ≥ 0.0075.
   - $y_{43}^D \geq 0.0005$.
   - 6 jets were reconstructed.
   - Jet with the largest b-discriminate is required to have $\mathcal{B}_{\rm Jet} > 0.3$.
2. The kinematic Neural Network, with 14 inputs, was required to have an output of ≥ 0.68.
3. The b-tagging Neural Network, with 15 inputs, was required to have an output of ≥ 0.92. The output distributions of the two networks are shown in Figure 5.15.

The numbers of events passing each requirement, compared with estimates from the background simulations normalized to the integrated luminosity of $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50$ pb^{-1}, are shown in Tables 5.17 and 5.18. Also shown are the detection efficiencies for various $e^+e^- \to h^0\,A^0 \to A^0\,A^0\,A^0 \to 3(b\bar{b})$ signal mass points. Figure 5.16 presents the signal efficiencies used to calculate the signal efficiency surface needed in order to perform the exclusion scan of the MSSM parameter space. Figure 5.17 presents the generated signal efficiency surface used in the scan.

After the final cut, 5 events were observed, consistent with the Standard Model expected background of $8.66 \pm 0.96 \pm 2.4$ events. Table 5.16 lists the Run and Event numbers of those events in the 189 GeV OPAL data which were selected by this analysis. Figure 5.18 shows a display event 53614 in run 10311, which had one of the highest ANN output values of the selected data events: 0.859 for the kinematic network; and 0.947 for the b-tagging network. The surviving events are compared to the the SM predicted distributions in Figures 5.19 - 5.26.

Data Events Selected

Run	Event
9644	33696
9646	66319
9840	6740
9937	25704
10311	53614

Table 5.16: Event numbers of those data events selected by the Neural Network based 189 GeV analysis.

ANN Analysis: Signal Results										
	Signal Mass points									
	(24,12)	(40,12)	(60,12)	(80,12)	(40,20)	(50,20)	(70,20)	(60,30)	(80,30)	(80,40)
Total	459	500	500	500	500	500	500	500	500	451
1. Preselection	317	280	334	392	318	293	452	436	459	425
2. Kinematic Net	206	199	243	315	276	251	413	419	409	380
3. B-Tagging Net	45	82	120	202	98	115	227	220	255	248
ϵ	.098	.164	.239	.405	.196	.230	.454	.440	.509	.550
$\pm$ (Stat)	.014	.016	.019	.022	.018	.019	.022	0.022	.022	.023
$\pm$ (Syst)	.030	.017	.037	.055	.048	.032	.043	.060	.080	.059

Table 5.17: The number of signal events surviving each cut for the 189 GeV data. Efficiencies for the signal events, denoted by the columns (m_{h^0}, m_{A^0}), are given in the last row along with the statistical and systematic error estimates.

ANN Analysis: Data and Background Results						
	Data	$(Z/\gamma)^* \to q\bar{q}$	$e^+e^- \to q\bar{q}q\bar{q}$	$e^+e^- \to q\ell q\ell$	$e^+e^- \to qeqe$	$\langle n \rangle$
Total	18701	16976.4 (148604)	1447.9 (27088)	1492.8 (43396)	4388.8 (99500)	24306.0
1. Preselection	1870	1205.0 (10548)	466.5 (8727)	46.8 (1359)	21.8 (493)	1739.91
2. Kinematic Net	649	438.33 (3837)	132.8 (2485)	14.0 (407)	5.8 (131)	590.95
3. B-Tagging Net	5	$7.7 \pm 0.9 \pm 2.0$ (67)	$0.9 \pm 0.2 \pm 0.5$ (16)	0.0 ± 0.0 (0)	0.1 ± 0.1(3)	$8.66 \pm 0.96 \pm 2.4$

Table 5.18: OPAL data compared to the number of predicted background Monte Carlo events surviving each cut and the expected background, $\langle N \rangle$, normalized to $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\,\mathrm{pb}^{-1}$. Bracketed values are the number of events passing the cuts from the Monte Carlo data sample. Errors are statistical and systematic, respectively.

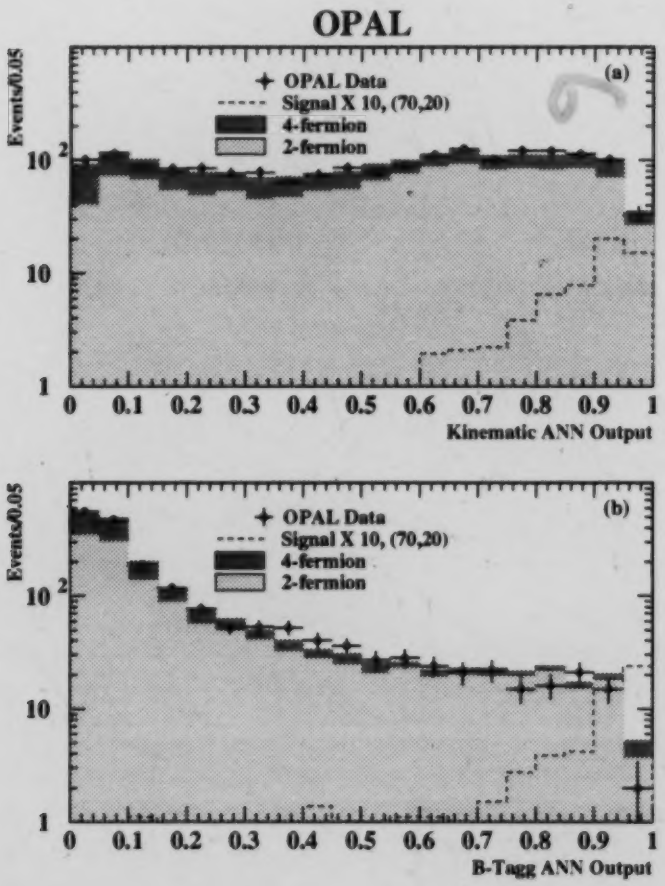


Figure 5.15: Presented are output distributions for the Kinematic (top) and B-tagging (bottom) neural networks used in this analysis. A simulated signal with (m_{h^0} , m_{A^0})=(70,20) GeV, with a representative $\sigma \times \mathrm{Br} = 0.04$ pb, is also shown.

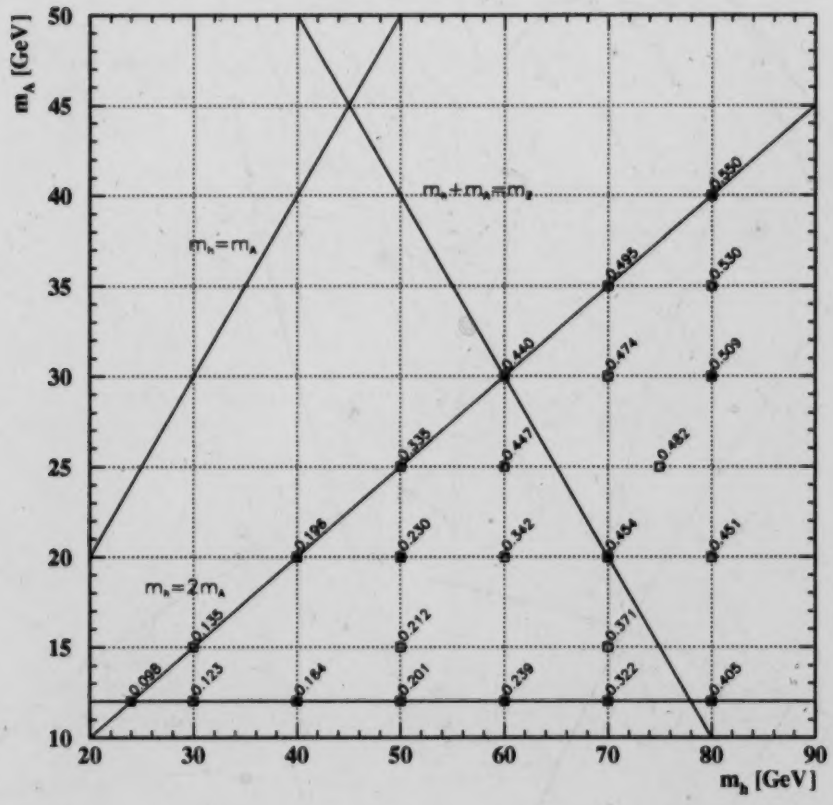


Figure 5.16: Signal efficiencies determined in the Neural Network based analysis. Dark squares are mass point efficiencies determined in this analysis. The open squares indicate interpolated efficiency points which were used in the generation of the efficiency surface, Figure 5.17, used in the MSSM scan.

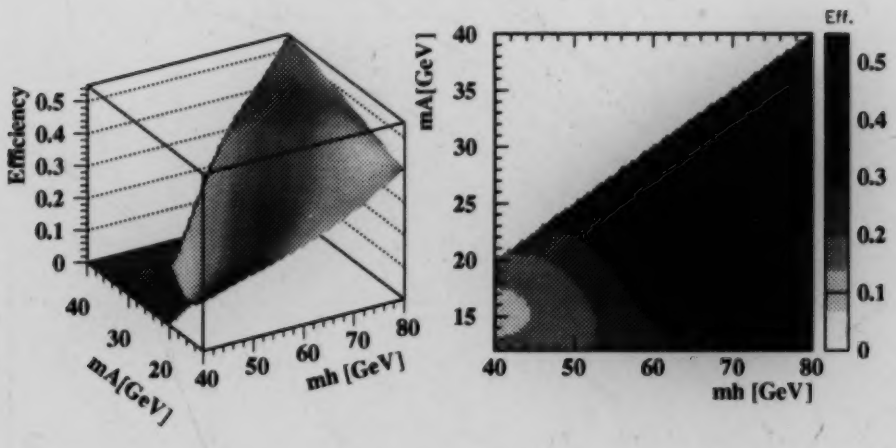


Figure 5.17: The signal efficiency surface used in the MSSM scan.

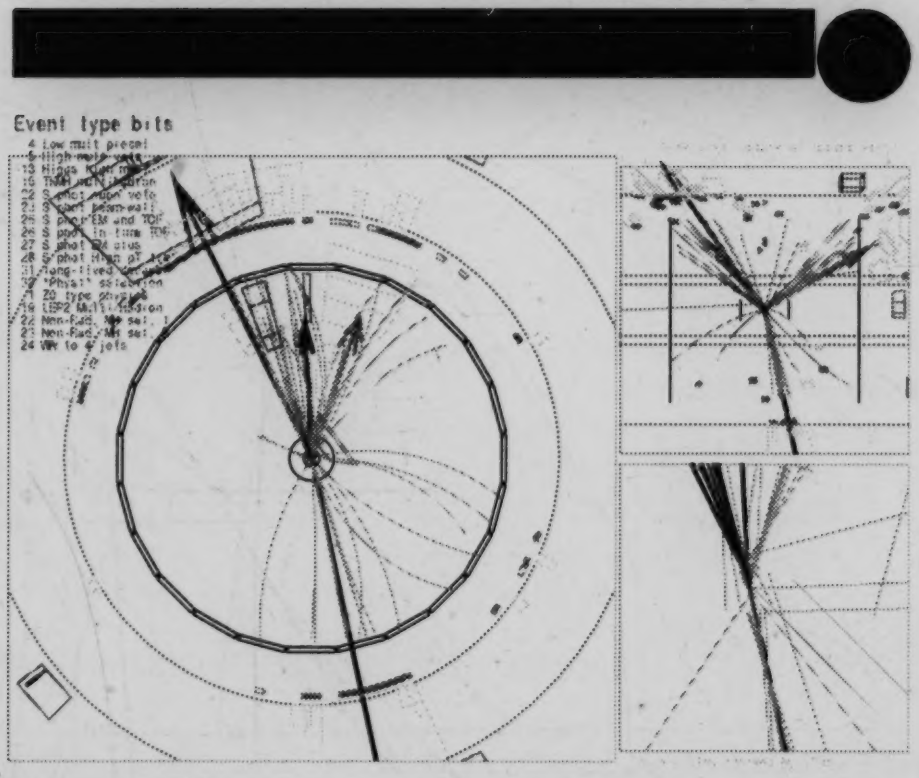


Figure 5.18: Event display of event 53614 in run 10311, selected by the 189 GeV neural network analysis as a signal candidate.

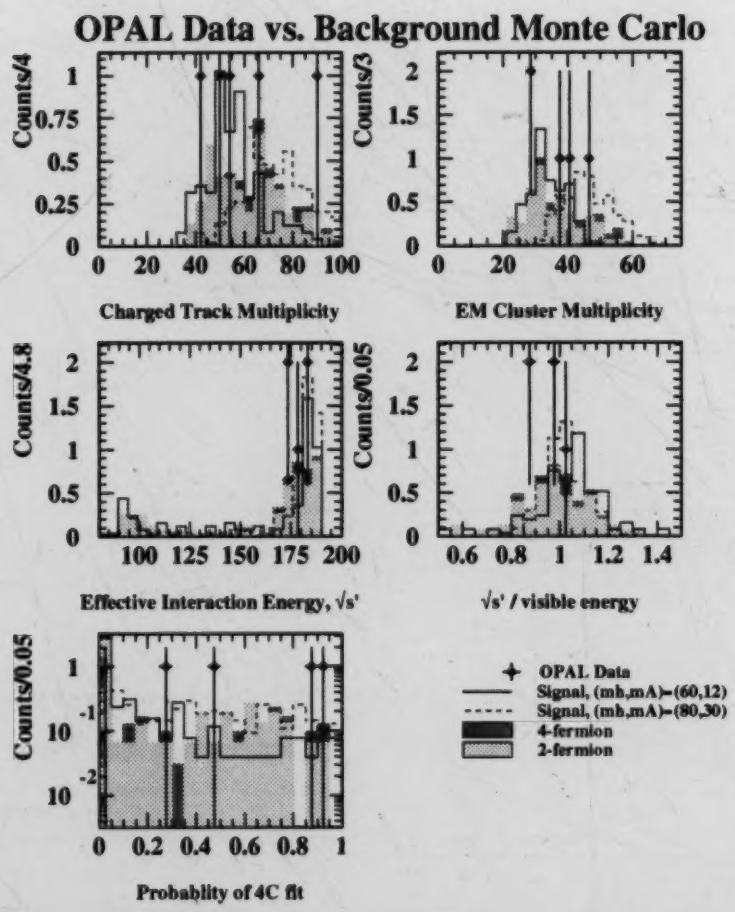


Figure 5.19: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

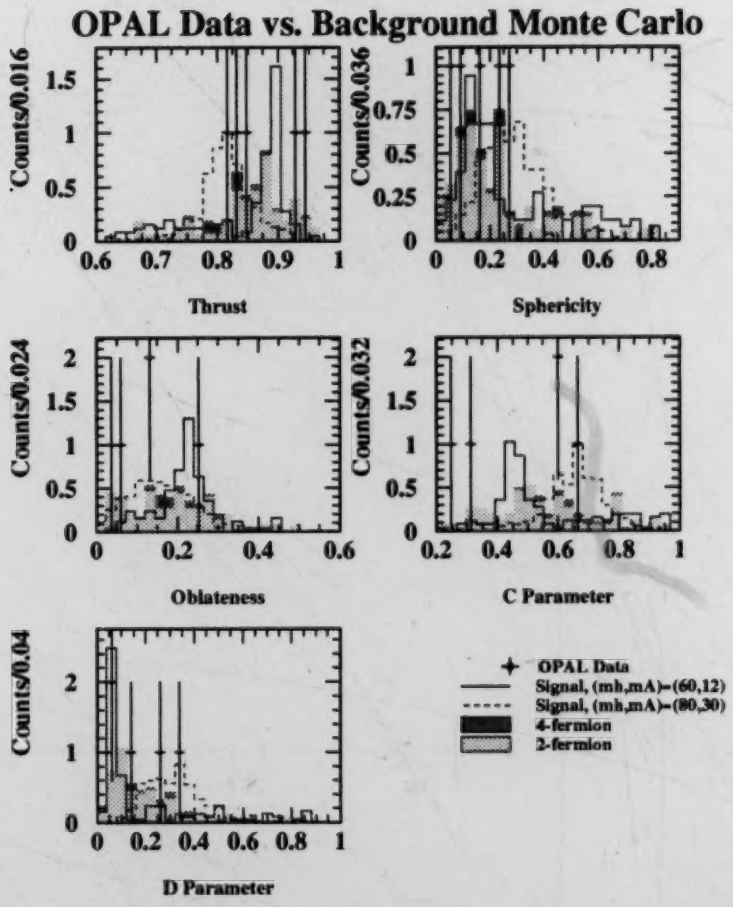


Figure 5.20: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

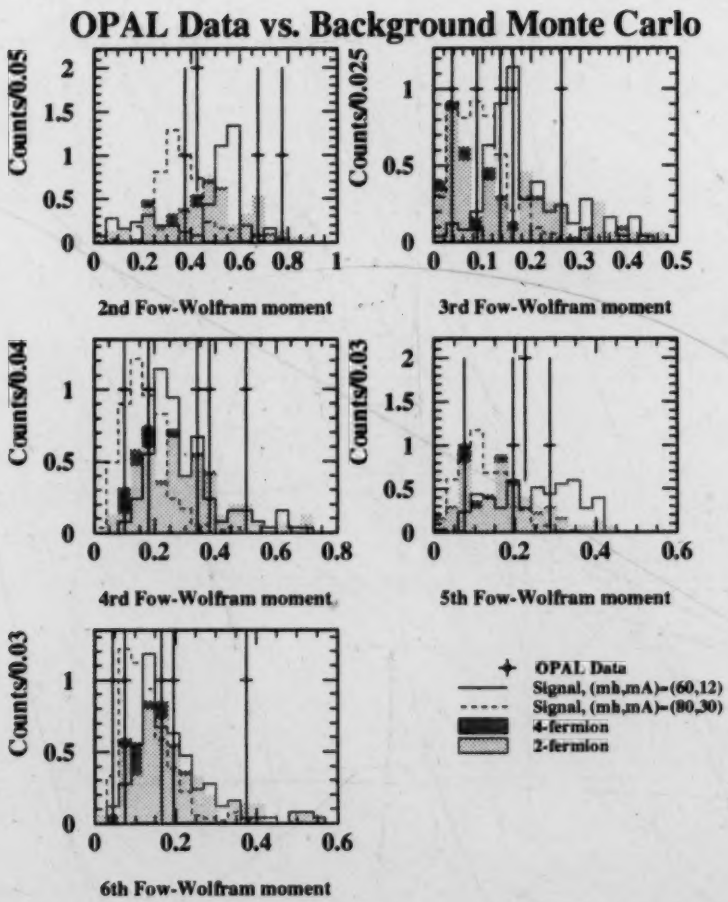


Figure 5.21: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

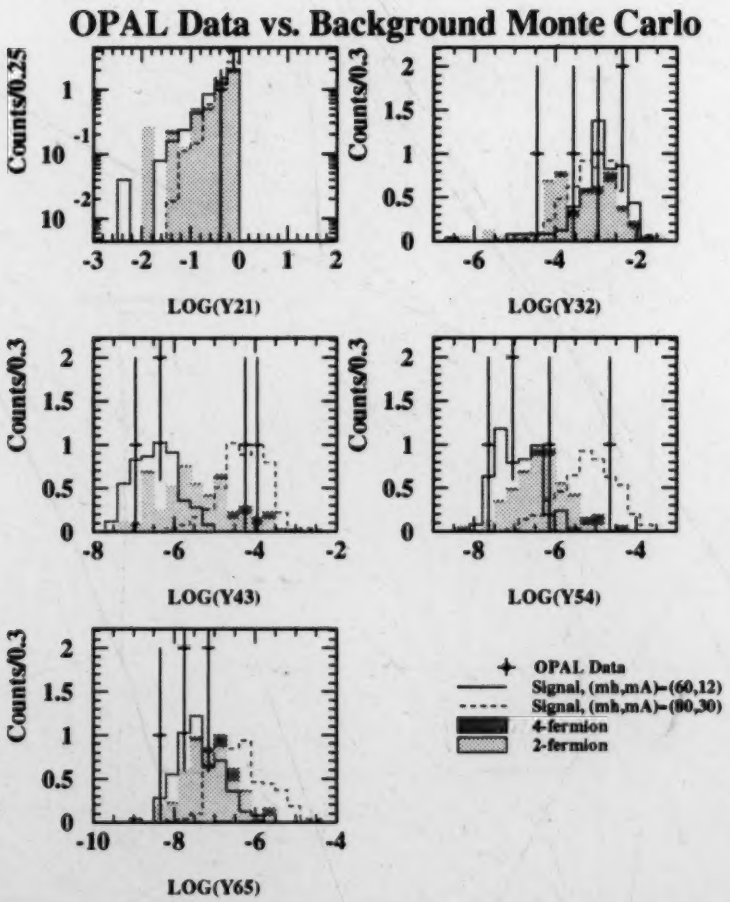


Figure 5.22: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

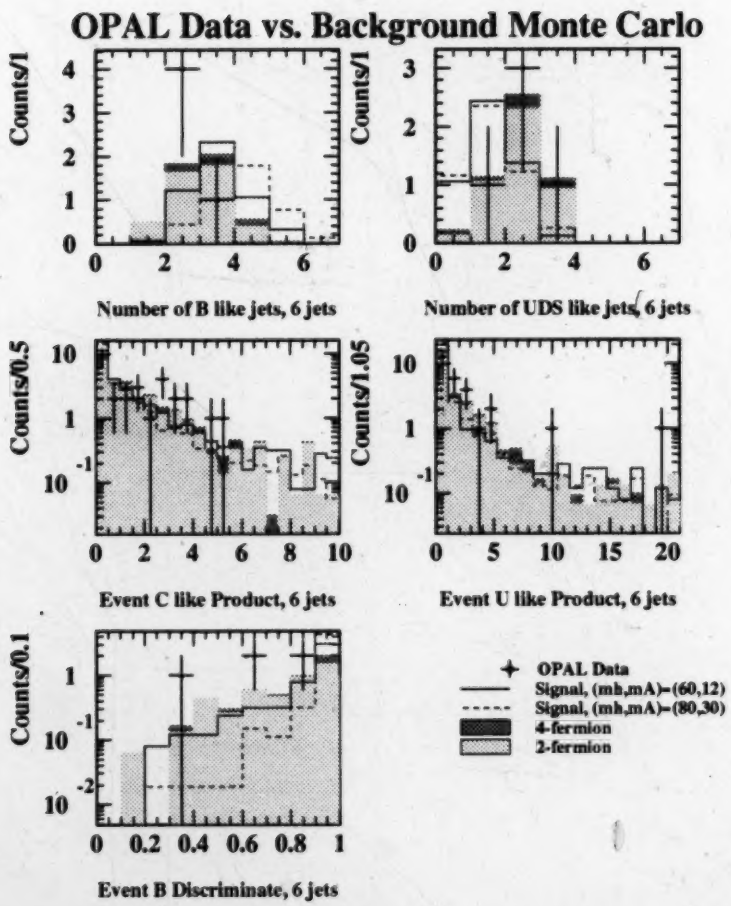


Figure 5.23: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

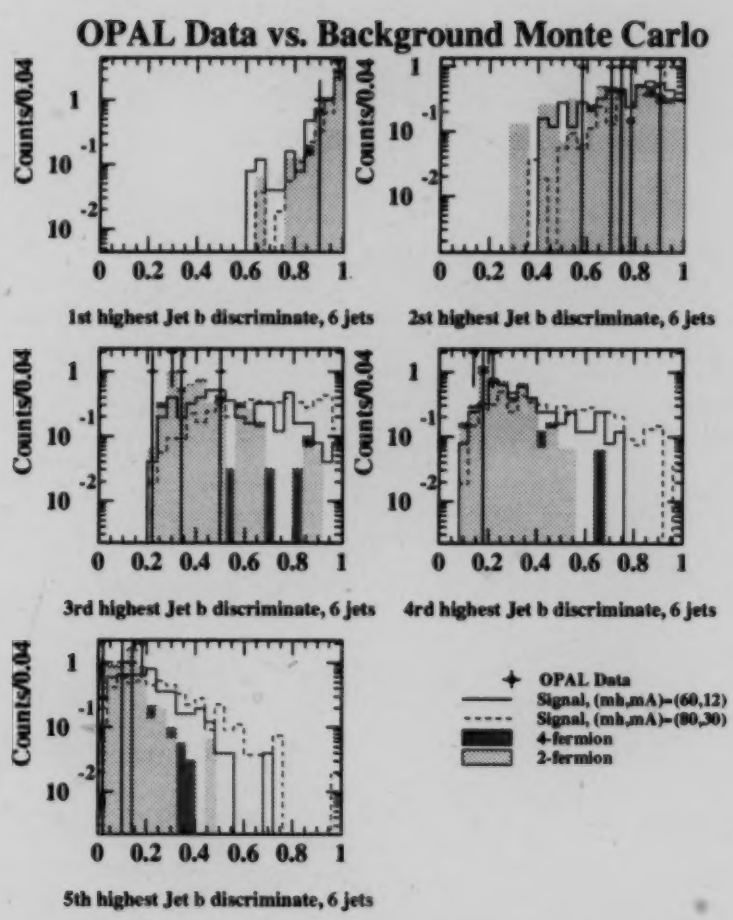

OPAL Data vs. Background Monte Carlo
Counts/0.04
1st highest Jet b discriminate, 6 jets
2st highest Jet b discriminate, 6 jets
3rd highest Jet b discriminate, 6 jets
4rd highest Jet b discriminate, 6 jets
5th highest Jet b discriminate, 6 jets
OPAL Data
Signal, (mh,mA)=(60,12)
Signal, (mh,mA)=(80,30)
4-fermion
2-fermion

Figure 5.24: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

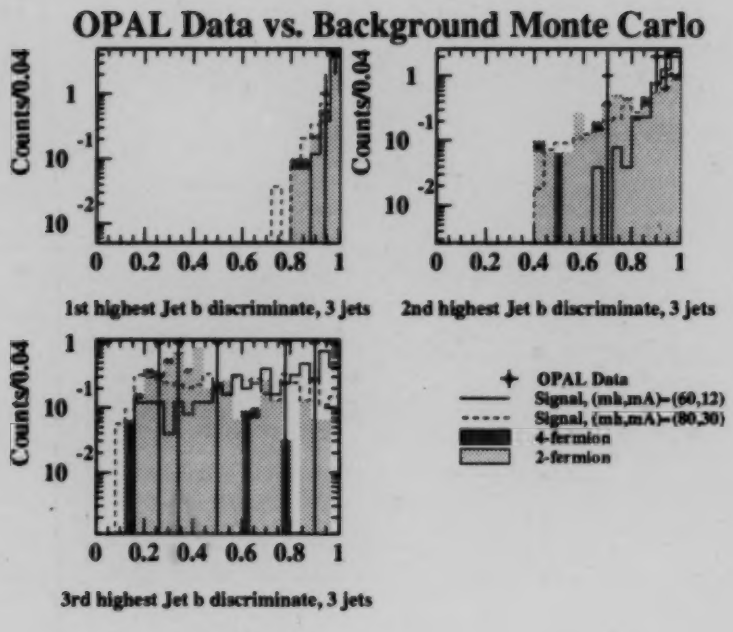


Figure 5.25: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

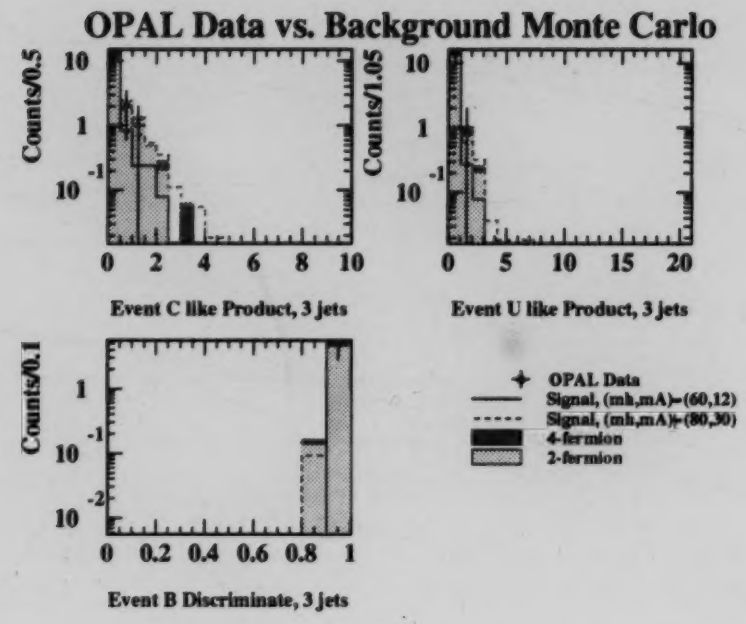


Figure 5.26: Distributions of the relevant analysis parameters for OPAL data (full circles) and the Standard Model prediction (solid histograms) after the analysis has been applied. Sample signal distributions at two mass points are shown (solid and dashed histogram), normalized to the data for ease of comparison.

5.6.1 MSSM Exclusion Limit

The exclusion limits from the $\sqrt{s} = 183\,\text{GeV}$ analysis were combined with this new scan of the MSSM parameter space (cf. section 1.2.2), using results from all OPAL Higgs analyses performed at $\sqrt{s} = 189\,\text{GeV}$ [113], to create the 95% CL exclusion limits shown in Figure 5.27. The black areas are excluded solely by comparing data to the expected MSSM production rates, without applying any additional theoretical criteria. The non-hatched grey areas are theoretically inaccessible.

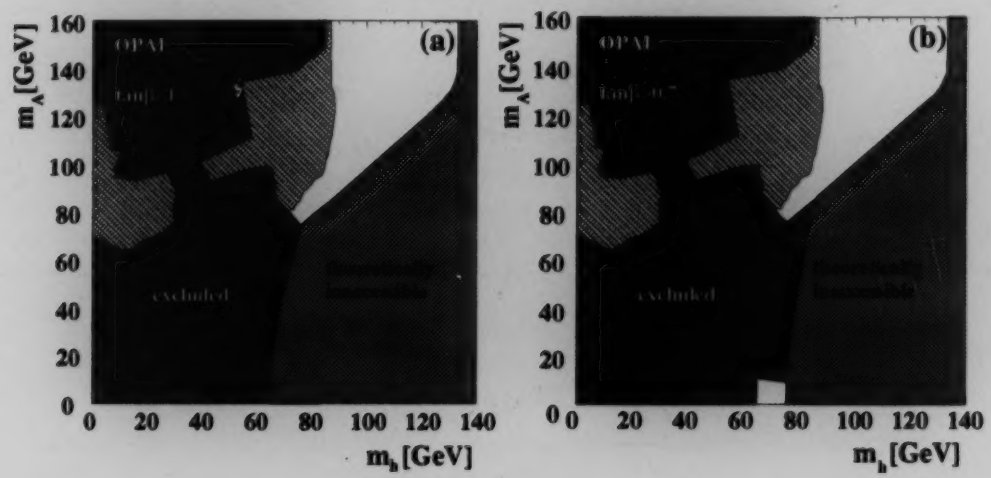


Figure 5.27: The MSSM exclusion for the 189 GeV analysis. Excluded regions are shown for (a) the (m_{h^0}, m_{A^0}) plane for $\tan\beta > 1$, (b) the (m_{h^0}, m_{A^0}) plane for $\tan\beta > 0.7$. All exclusion limits are at 95% CL. The dark grey areas are excluded solely by comparing data to the expected MSSM production rates, without applying any additional theoretical criteria. The unhatched light greay areas are theoretically inaccessible.

The theoretically inaccessible limit defined by $m_{h^0} \approx 80\,\text{GeV}$ and $0\,\text{GeV} \leq m_{A^0} \lesssim 80\,\text{GeV}$ was shifted, compared to the previous analysis, to slightly higher m_{h^0} values due to the higher centre of mass energy. High signal efficiencies from the analysis presented in this chapter, for mass points which lie in this border region, permitted the exclusion limit to again buttress the theoretically inaccessible area. The lower limits on m_{h^0} and m_{A^0} slightly improved over the previous result as the unexcluded, 5 GeV $< m_{A^0} <$ 15 GeV and 65 GeV $< m_{h^0} <$ 72 GeV, observed in the $\sqrt{s} = 183\,\text{GeV}$ limit was reduced to the $A^0 \rightarrow b\bar{b}$ threshold, $m_{A^0} < 10\,\text{GeV}$. An unexcluded region still remains due to the insensitivity of this analysis to signals with $m_{A^0} < 10$ GeV. In the MSSM, we obtain the absolute mass limits of $m_{h^0} > 72.2$ GeV and $m_{A^0} > 76.0$ GeV for $\tan\beta > 1$.

Chapter 6

Conclusion

This thesis presented Neural Network based searches for $e^+e^- \rightarrow Z^0 \rightarrow h^0A^0 \rightarrow A^0A^0A^0$ with $A^0 \rightarrow b\bar{b}$ using data collected by the OPAL detector at $\sqrt{s} = 183\,\text{GeV}$, with a integrated luminosity of $\mathcal{L} = 54.5 \pm 0.2\ \text{pb}^{-1}$, and 189 GeV, with $\mathcal{L} = 171.09 \pm 0.15 \pm 0.50\ \text{pb}^{-1}$. Exclusion limits were calculated in the constrained Minimal Supersymmetric Model and Higgs production through this channel was excluded, at 95% CL, for $m_{h^0} \geq 2m_{A^0}$, $m_{h^0} \lesssim 80GeV$, $m_{A^0} > 12\,\text{GeV}$.

As there was no evidence for a Higgs signal with $m_{h^0} \geq 2m_{A^0}$ in the mass regions considered, more general limit on the 2HDM process $e^+e^- \rightarrow S^0P^0 \rightarrow P^0P^0P^0$ at $\sqrt{s} = 189\,\text{GeV}$, where S^0 and P^0 denote scalar and pseudo-scalar neutral bosons, respectively, can be drawn by identifying S^0 and P^0 with h^0 and A^0. The $\sqrt{s} = 189\,\text{GeV}$ analysis presumed that the Higgs bosons in $h^0A^0 \rightarrow A^0A^0A^0$ decayed into b-quark pairs, thus, only the upper limit for the combined cross-section times branching ratio, σ_{BR}, can be calculated. Figure 6.1 presents the σ_{BR} exclusion surface, at 90% and 95% confidence level, for the results of the 189 GeV neural network analysis calculated using the method described in Appendix A. The maximal production rate, at 95% CL, for $h^0A^0 \rightarrow A^0A^0A^0 \rightarrow 3(b\bar{b})$ admitted by the data ranges from $\sigma_{BR} \leq 0.192$ pb for $(m_{h^0}, m_{A^0}) = (40, 12)\,\text{GeV}$ to $\sigma_{BR} \leq 0.057$ pb for $(m_{h^0}, m_{A^0}) = (80, 40)\,\text{GeV}$.

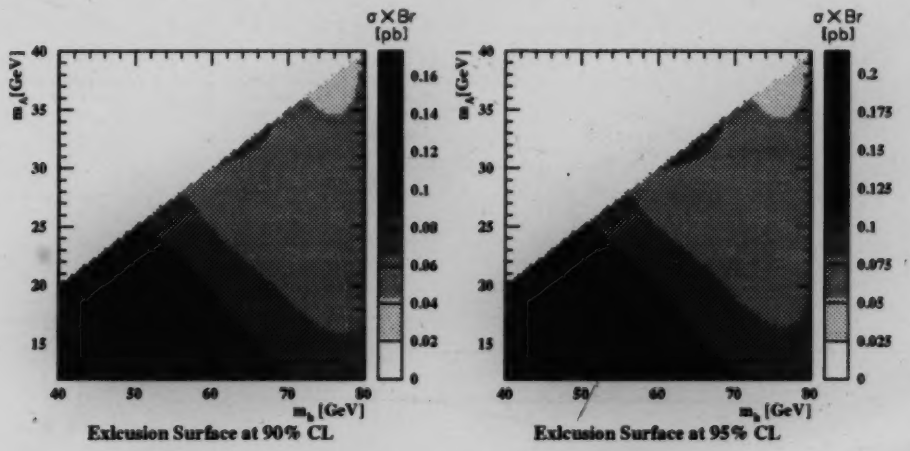


Figure 6.1: The 2HDM exclusion surface derived from the 189 GeV neural network analysis. The shaded contour plots indicate the maximum cross-section times branching ratio allowed by the data over the relevant mass region.

Bibliography

[1] M. Gell-Mann, "Nonleptonic weak decays and the eightfold way," *Phys. Rev. Lett.* **12** (1964) 155–156. ; G. Zweig, "An su(3) model for strong interaction symmetry and its breaking. 2,". CERN-TH-412.

[2] S. Glashow, "Partial symmetries of weak interactions," *Nucl. Phys.* **22** (1961) 579; A. Salam, "Weak and electromagnetic interactions," in *Elementary Particle Theory*, W. Svartholm, ed., p. 367. Almquist and Wiksell, Stockholm, 1968; S. Weinberg, "A model of leptons," *Phys. Rev. Lett.* **19** (1967) 1264.

[3] D.J. Gross and F. Wilczek *Phys. Rev.* **D8** (1973) 3633; D.J. Gross and F. Wilczek *Phys. Rev.* **D9** (1974) 908; H.D. Politzer *Phys. Rev.* **C14** (1974) 129; F. Wilczek *Ann. Rev. Nucl. Part. Sci.* **32** (1982) 177.

[4] For example: **LEP** Collaboration, D. Abbaneo *et. al.*, "A Combination of preliminary Electroweak measurements and constraints on the Standard Model,". CERN-EP-99-015; D. Karlen, "Experimental status of the standard model," in *XXIX International Conference on High Energy Physics*, ICHEP98. Vancouver, BC, 1998. To be published.

[5] S. Dawson, "Introduction to the physics of Higgs bosons," `hep-ph/9411325`. ; J. L. Hewett, "The Standard model and why we believe it," `hep-ph/9810316`. .

[6] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions*. Benjamin-Cummings, Reading, MA, 1983; L. Ryder, *Quantum Field Theory*. Cambridge University Press, New York, NY, 2 ed., 1996.

[7] P.W. Higgs *Phys. Rev.* **12** (1964) 132; F. Englert and R. Brout *Phys. Rev. Lett.* **13** (1964) 321; T.W.B. Kibble *Phys. Rev.* **D155** (1967) 1554; G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble *Phys. Rev. Lett.* **13** (1956) 585.

[8] G. 't Hooft *Nucl. Phys.* **B35** (1971) 167; G. 't Hooft and M.T. Veltman *Nucl. Phys.* **B44** (1973) 189; B.W. Lee and J. Zinn-Justin *Phys. Rev.* **D5** (1972) 3121.

[9] E. Abers and B. Lee *Phys. Rev.* **9** (1975) 1.

[10] **Super-Kamiokande** Collaboration, T. Kajita *et. al.*, "Atmospheric neutrino results from super Kamiokande and Kamiokande: Evidence for neutrino(mu) oscillations," `hep-ex/9810001`. .

[11] N. Cabibbo *Phys. Rev. Lett.* **10** (1963) 531; M. Kobayashi and T. Maskawa *Prog. Theor. Phys.* **49** (1973) 652.

[12] **GARGAMELLE** Collaboration, F. Hasert *et. al. Phys. Lett.* **B46** (1973) 138.

[13] **UA1** Collaboration, G. Arnison *et. al. Phys. Lett.* **B122** (1983) 103; E. Braaten and J. Leveille *Phys. Rev.* **D22** (1980) 715; M. Drees and K. Hikasa *Phys. Lett.* **B240** (1990) 455.

[14] A. Djouadi and P. Gambino *Phys. Rev. Lett.* **73** (1994) 2529.

[15] B. Ioffe and V. Khoze *Sov. J. Part. Nucl. Phys.* **9** (1978) 50.

[16] M. Carena and P. Zerwas, "Higgs Physics," in *Physics at LEP2* [101].

[17] **ALEPH** Collaboration, "Limits on Higgs boson masses from combining the data of the four LEP experiments at $\sqrt{s} \leq 183$ GeV,". CERN-EP-99-060.

[18] S. Weinberg *Phys. Rev.* **D13** (1976) 974; S. Weinberg *Phys. Rev.* **D19** (1979) 1277; L. Susskind *Phys. Rev.* **D20** (1979) 2619.

[19] S. P. Martin, "A Supersymmetry primer," **hep-ph/9709356**. .

[20] E. Witten *Nucl. Phys.* **B188** (1981) 513; N. Sakai *Z. Phys.* **C11** (1981) 153; S. Dimopoulos and H. Georgi *Nucl. Phys.* **B193** (1981) 150.

[21] Y. Gol'fand and E. Likhtman *JETP Lett.* **13** (1971) 323; D. Volkov and V. Akulov *Phys. Lett.* (1973) 323.

[22] J. Wess and B. Zumino *Nucl. Phys.* **B70** (1974) 39.

[23] H. Haber and Y. Nir *Nucl. Phys.* **B335** (1990) 363.

[24] H. Nilles *Phys. Rev.* **110** (1984) 1; H. Haber and G. Kane *Phys. Rev.* **117** (1985) 75; R. Barbieri *et. al.*, "Z Physics at LEP1," CERN Technical Report CERN 89-08, CERN, 1984. Vol. 2, p. 121; J.M. Frère, G.L. Kane *Nucl. Phys.* **B223** (1983) 331; J. Ellis *et. al. Phys. Lett.* **B123** (1983) 436; J. Ellis *et. al. Phys. Lett.* **B127** (1983) 233.

[25] J.F. Gunion, H.E. Haber, G.L. Kane and S. Dawson, *The Higgs Hunters Guide.* Addison-Wesley Publishing Company, (Reading, MA), 1990.

[26] **CDF** Collaboration, F. Abe *et. al. Phys. Rev. Lett.* **74** (1995) 2626; **D0** Collaboration, S. Abachi *et. al. Phys. Rev. Lett.* **74** (1995) 2632.

[27] H. Haber and R. Hempfling *Phys. Rev. Lett.* **66** (1991) 1815; J. Ellis, G. Ridolfi and F. Zwirner *Phys. Lett.* **B257** (1991) 83; M. Berger *Phys. Rev.* **D41** (1990) 225; M. Y. Y. Okada and T. Yanagida *Prog. Theor. Phys. Lett.* **85** (1991) 1; M. Carena, M. Quiros and C. Wagner *Nucl. Phys.* **B461** (1996) 407.

[28] **OPAL** Collaboration, K. Ackerstaff *et. al.*, "A Search for neutral Higgs bosons in the MSSM and models with two scalar field doublets," *Eur. Phys. J.* **C5** (1998) 19, **hep-ex/9803019**. .

[29] P. Janot, "HZHA Higgs event generator. Version 2.00," in *Physics at LEP2* [101], p. 309.

[30] A. Djouadi, J. Kalinowski and M. Spira, "**HDECAY**: a program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension," DESY Preprint DESY-97-079, DESY, Apr., 1997.

[31] M. L. Mangano *et. al.*, "Event generators for discovery physics," **hep-ph/9602203**. .

[32] "LEP Design Report," CERN Technical Report CERN-LEP/84-01, CERN, 1984.

[33] M. Martinez, R. Miquel, L. Rolandi, and R. Tenchini *Rev. Mod. Phys.* **71** (1999) 575. .

[34] **OPAL** Collaboration, "OPAL Technical Proposal," CERN Technical Report CERN/LEPC/83-4, CERN, 1990.

[35] **OPAL** Collaboration, K. Ahmet *et. al. Nucl. Instrum. Methods* **A305** (1991) 275–319.

[36] **ALEPH** Collaboration, D. Decamp *et. al. Nucl. Instrum. Methods* **A294** (1990) 121.

[37] **DELPHI** Collaboration, P. Aarnio *et. al. Nucl. Instrum. Methods* **A303** (1991) 233.

[38] **L3** Collaboration, B. Adeva *et. al. Nucl. Instrum. Methods* **A289** (1990) 35.

[39] J. M. Jowett, "Beam dynamics at LEP,". Invited talk at 14th Advanced ICFA Beam Dynamics Workshop: Beam Dynamics Issues for e^+e^- Factories (ICFA 97), Frascati, Italy, 20-26 Oct 1997.

[40] H. Breuker *et. al. Nucl. Instrum. Methods* **A260** (1987) 329.

[41] **OPAL** Collaboration, P. Allport *et. al. Nucl. Instrum. Methods* **A324** (1993) 34.

[42] **OPAL** Collaboration, P. Allport *et. al. Nucl. Instrum. Methods* **A346** (1994) 476.

[43] **OPAL** Collaboration, G. Alexander *et. al. Phys. Lett.* **B374** (1996) 341.

[44] **OPAL** Collaboration, K. Akerstaff *et. al. Z. Phys.* **C73** (1997) 397; **OPAL** Collaboration, K. Akers *et. al. Z. Phys.* **C68** (1995) 531; **OPAL** Collaboration, K. Akers *et. al. Phys. Lett.* **B350** (1995) 273; **OPAL** Collaboration, K. Akers *et. al. Z. Phys.* **C67** (1995) 379.

[45] S. Anderson *et. al. Nucl. Instrum. Meth.* **A403** (1998) 326. .

[46] J. Carter *et. al. Nucl. Instrum. Methods* **A286** (1990) 99.

[47] A. Carter *et. al. Nucl. Instrum. Methods* **A286** (1990) 107.

[48] R. Heuer *et. al. Nucl. Instrum. Methods* **A265** (1988) 11.

[49] H. Fischer *et. al. Nucl. Instrum. Methods* **A283** (1986) 331.

[50] H. Fischer *et. al. Nucl. Instrum. Methods* **A283** (1988) 492.

[51] O. Biebel *et. al. Nucl. Instrum. Methods* **A323** (1992) 169.

[52] H. Mes *et. al. Nucl. Instrum. Methods* **A323** (1992) 169.

[53] R. Kowalewski, "Study of jet finding for b physics in OPAL," OPAL Technical Report OPAL-TN/180, The OPAL Experiment, CERN.

[54] L. del Pozo, "Jet Finding with a Cone Algorithm," OPAL Technical Report OPAL-TN/170, The OPAL Experiment, CERN, 1993.

[55] G. Aguillion *et. al.*, "Thin scintillating tiles with high light yield for the OPAL endcaps," *Nucl. Instrum. Methods* **A417** (1998) 266–277.

[56] P. Mattig, "Photon Energy Resolution in the Lead-glass Calorimeters," OPAL Technical Report OPAL-TN/324, The OPAL Experiment, CERN, Oct., 1995.

[57] **OPAL** Collaboration *Nucl. Instrum. Methods* **A313** (1992) 103.

[58] K. Gurney, *An Introduction to Neural Networks*. UCL Press, 1 Gunpowder Square, London, 1 ed., 1996; R. Davis, "Search for a Standard Model Higgs Boson on 161 GeV Electron Positron Collisions using Neural networks," Master's thesis, University of Alberta, Edmonton, Alberta, Spring, 1997.

[59] S. J. Heims, *John von Neumann and Norbert Weiner - From Mathematics to the Technologies of Life and Death*. Academic Press, 1982.

[60] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics* **7** (1943) 115–133.

[61] D. O. Hebb, *The Organization of Behaviour*. John Wiley, 1949.

[62] F. Rosenblatt, *Principles of Neurodynamics*. Spartan Books, 1962.

[63] M. Minsky and S. Papert, *Perceptrons*. MIT Press, 1969.

[64] D.E. Rumelhart, G.E. Hinton and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323** (1986) 533–536.

[65] B. Denby, "Neural networks and cellular automata in experimental high- energy physics," *Comput. Phys. Commun.* **49** (1988) 429. .

[66] C. Peterson, "Track finding with neural networks," *Nucl. Instr. Meth.* **A279** (1989) 537. .

[67] B. Denby, "The use of neural networks in high-energy physics,". FERMILAB-PUB-92-215-E; H. Kolanoski, "Application of artificial neural networks in particle physics," *Nucl. Instrum. Meth.* **A367** (1995) 14–20. .

[68] L. Lonnblad, C. Peterson, and T. Rognvaldsson, "Finding gluon jets with a neural trigger," *Phys. Rev. Lett.* **65** (1990) 1321–1324. ; I. Csabai, F. Czako, and Z. Fodor, "Quark and gluon jet separation using neural networks," *Phys. Rev.* **D44** (1991) 1905–1908. ; M. A. Graham, L. M. Jones, and S. Herbin, "A neural network classification of quark and gluon jets," *Phys. Rev.* **D51** (1995) 4789–4807. .

[69] J. Proriol, A. Falvard, P. Henrard, J. Jousset, and B. Brandl, "Tagging b and c quark events in e+ e- collisions with neural networks,". In *Marciana Marina 1992, Neural networks* 267-270; K. H. Becks, F. Block, J. Drees, P. Langefeld, and F. Seidel, "B quark tagging using neural networks and multivariate statistical methods: A comparison of both techniques," *Nucl. Instrum. Meth.* **A329** (1993) 501–517. ; **D0** Collaboration, B. Abbott *et. al.*, "Neural networks for analysis of top quark production," **hep-ex/9907041**. .

[70] L. Ametller, L. Garrido, G. Stimpfl-Abele, P. Talavera, and P. Yepes, "Discriminating signal from background using neural networks. application to top quark search at the fermilab tevatron," *Phys. Rev.* **D54** (1996) 1233–1236, **hep-ph/9603269**. ; T.Maggipinto, G.Nardulli, S.Dusini, F.Ferrari, I.Lazzizzera, A.Sidoti, A.Sartori, and G.P.Tecchiolli, "Role of neural networks in the search of the higgs boson at LHC," *Phys. Lett.* **B409** (1997) 517–522, **hep-ex/9705020**.

[71] G. Alimonti and S. Magni, "A neural network approach to spatial reconstruction in the ctf detector," *Nucl. Instrum. Meth.* **A411** (1998) 467. .

[72] C. S. Lindsey, B. Denby, H. Haggerty, and K. Johns, "Drift chamber tracking with a vlsi neural network,". 2nd Workshop on Neural Networks: from Biology to High Energy Physics, Marciana Marina, Italy, 18-26 Jun 1992; C. S. Lindsey, B. Denby, and H. Haggerty, "Drift chamber tracking with neural networks," *IEEE Trans. Nucl. Sci.* **40** (1993) 607–614. .

[73] R. Nobrega and J. Varela, "Neural networks in hep triggers,". Talk given at Computing in High-energy Physics (CHEP 97), Berlin, Germany, 7-11 Apr 1997.

[74] C. Peterson, T. Rögnvaldsson and L. Lönnblad, "`JETNET 3.0` - A Versatile Artificial Neural Network Package," Software Manual LT TP 93-29, CERN-TH.7135/94, Department of Theoretical Physics, University of Lund, Dec., 1993; M.D. Richard and R.P. Lippmann *Neural Comput.* **3** (1991) 461.

[75] B.J.A. Krose and P. van der Smagt, "An Introduction to Neural Networks," Course Notes, University of Amsterdam, Jan., 1993.

[76] E. Baum and D. Haussler *Neural Comput.* **1** (1989), no. 1, 151.

[77] R. Hawkings, "Vertex tagging of b-hadrons using 3D tracking and neural networks," OPAL Technical Note OPAL-TN366, CERN, 1996; G. Bahan and R. Barlow, "Identification of b jets using neural networks," OPAL Institute Publication MAN/HEP/92/1, University of Manchester, 1992.

[78] R. Ibata. astro-ph/9704014, Apr., 1997.

[79] For example: S. .Ronen, A. Aragón-Salamanca, and O. Lahav *Mon. Not. R. Astron. Soc.* (May, 1998) 11, `astro-ph/9805130`; M. Turler and T. Courvoisier *Mon. Not. R. Astron. Soc.* (Sept., 1997) 11, `astro-ph/9709038`; Z. Bagoly *et. al.* astro-ph/9708084, Aug., 1997; K. Glazebrook, A. Offer, and K. Deeley. astro-ph/9707140, 1997.

[80] The Particle Data Group, L. Montanet *et. al.*, *Eur. Phys. J., Review of Particle Physics*, vol. 3. Springer-Verlag, Secaucus, NJ, 1998.

[81] The Numerical Algorithms Group, Inc., "The NAG Library." NAG Ltd., Oxford, UK. http://www.nag.co.uk/.

[82] R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences.* John Wiley & Sons, Ltd., Baffins Lane (Chichester), 1989.

[83] T. Sjöstrand, "PYTHIA 5.7 and JETSET 7.4: Physics and Manual." LU-TP/95-20, CERN-TH.7112/93, Aug., 1995; T. Sjöstrand *Comp. Phys. Comm.* **39** (1986) 347; T. Sjöstrand and M. Bengtsson *Comp. Phys. Comm.* **43** (1987) 367; T. Sjöstrand and M. Bengtsson *Comp. Phys. Comm.* **82** (1994) 74.

[84] S. Kawabata *et. al.*, "GRC4F: User-Guide Version 2.1 Preliminary. GRACE manual: Automatic generation of tree amplitudes in Standard Models: Version 1.0." KEK-92-19, 1992; J. Fujimoto *et. al.*, "GRC4F v1.1: A four-fermion event generator for e^+e^- collisions," *Comp. Phys. Comm.* **100** (1997) 74.

[85] G. Marchesini *et. al. Comp. Phys. Comm.* **67** (1992) 465.

[86] J. Allison *et. al. Nucl. Instrum. Methods* **A317** (1992) 47.

[87] R. Bately, R. Hawkings, M. Hildreth, "`ODSIZL 2.11`, DST-Level Tracking Adjustments and Refit with Silicon Hits," OPAL Technical Report OPAL-TN/520, The OPAL Experiment, CERN, 1997.

[88] "Precision Luminosity for Z^0 Lineshape Measurements with the OPAL Silicon-Tungsten Luminometer," OPAL Technical Report OPAL-TN/528, The OPAL Experiment, CERN, 1997.

[89] S. Lloyd *et. al.*, "The OPAL Primer." An OPAL Experiment reference manual, July, 1997.

[90] S. Yamashita, "`DH` (Discover Higgs) Ntuple for 4-jet Final State at LEP-II. Version 2.10." OPAL Higgs Group internal note.

[91] P.Igo-Kemenes, U.Jost, G.Pasztor and S.Yamashita, "Search in the four-jet topology for heavy particles, pair-produced in e^+e^- collisions at 161 GeV," OPAL Physics Note OPAL-PN/235, The OPAL Experiment, CERN, 1996.

[92] T.Omori, S.Asai, I.Nakamura and S.Yamashita, "A Matching Algorithm: MT package," OPAL Technical Report OPAL-TN/381, The OPAL Experiment, CERN, 1996.

[93] G. Hanson *et. al.*, "Evidence for jet structure in hadron production by e+ e- annihilation," *Phys. Rev. Lett.* **35** (1975) 1609. .

[94] G. Sterman and S. Weinberg *Phys. Rev. Lett.* **39** (1977) 1436.

[95] **JADE** Collaboration, S. Bethke *et. al. Phys. Lett.* **B213** (1988) 235.

[96] N. Brown and W. J. Stirling *Phys. Lett.* **B252** (1990) 657; S. Catani *et. al. Z. Phys.* **B** (1991) 432; D. S. S. Bethke, Z. Kunszt and W. J. Stirling *Nucl. Phys.* **B370** (1992); R.K. Ellis, W.J. Stirling and B.R. Webber, *QCD and Collider Physics*. Cambridge University Press, Cambridge, UK, 1996.

[97] C.P. Ward, D.R. Ward and C. Hartmann, "`WW` Manual." An OPAL Experiment program manual.

[98] J. Bjorken and S. Brodsky *Phys. Rev.* **D1** (1970) 1416.

[99] S. Brandt, C. Peyrou, R. Sosnowski, and A. Wroblewski *Phys. Rev. Lett.* **12** (1964) 57; E. Fahri *Phys. Rev. Lett.* **39** (1977) 1587.

[100] G. Fox and S. Wolfram *Nucl. Phys.* **B149** (1979); G. Gary, "An OPAL Physics Utility Library PX Library, version 1.11." An OPAL Experiment program manual, 1997.

[101] G. Altarelli, T. Sjöstrand and F. Zwirner, ed., *Physics at LEP2*, vol. 2. CERN, Feb., 1996.

[102] **OPAL** Collaboration, G. Alexander *et. al. Phys. Lett.* **B376** (1996) 232–244.

[103] V.D. Barger and R.J.N. Phillips, *Frontiers in Physics: Collider Physics*. Addison-Wesley, Don Mills (Ontario), 1987.

[104] **OPAL** Collaboration, G. Abbiendi *et. al.*, "Search for Higgs bosons in e^+e^- collisions at 183-GeV," *Eur. Phys. J.* **C7** (1999) 407, `hep-ex/9811025`. .

[105] **OPAL** Collaboration, R. Akers *et. al. Z. Phys.* **C63** (1994) 197.

[106] **OPAL** Collaboration, R. Akers *et. al. Z. Phys.* **C66** (1995) 19; **OPAL** Collaboration, K. Ackerstaff *et. al. Z. Phys.* **C74** (1997) 1–17.

[107] R. Hawkings, "High performance b-tagging using 3d vertexing and neural networks," OPAL Technical Report OPAL-TN/515, The OPAL Experiment, CERN, Sept., 1997.

[108] "Search for neutral Higgs Bosons of the MSSM in e^+e^- collisions at $\sqrt{s}$ from 130 to 172 GeV," CERN Report CERN-PPE/97-071, The ALEPH Experiment, CERN, 1997.

[109] **OPAL** Collaboration, G. Alexander *et. al. Z. Phys.* **C70** (1996) 357.

[110] Giulio D'Agostini, "Bayesian Reasoning in High Energy Physics - Principles and Applications." CERN Lectures, Summer, 1998.

[111] P. Vikas *et. al.*, "`BT` code manual." An OPAL Experiment program manual.

[112] S. Baumann, M. Schumacher, E. von Törne, "`NT`: Analysis Ntuple for Higgs Search. Version 3.05." OPAL Higgs Group internal note, 1998.

[113] **OPAL** Collaboration, G. Abbiendi *et. al.*, "Search for Higgs Bosons in e^+e^- Collisions at 189 GeV," OPAL Physics Note, submitted to Eur. Phys. J. C OPAL-PN/382, The OPAL Experiment, CERN, May, 1999.

[114] **ALEPH** Collaboration, D. Buskulic *et. al. Phys. Lett.* **B357** (1995) 699; **ALEPH** Collaboration, D. Buskulic *et. al. Z. Phys.* **C62** (1994) 179; **DELPHI** Collaboration, P. Abreu *et. al. Z. Phys.* **C66** (1995) 323; **OPAL** Collaboration, R. Akers *et. al. Z. Phys.* **C60** (1993) 199; **OPAL** Collaboration, G. Alexander *et. al. Phys. Lett.* **B364** (1995) 93.

[115] G. Zech *Nucl. Instrum. Methods* **A277** (1989) 608–610.

[116] R. Cousins and V. Highland *Nucl. Instrum. Methods* **A320** (1992) 331.

Appendix A

Derivation of Exclusion Limit

In this Appendix, we will present a derivation of the calculation used to determine the upper limit on a hypothesized signal production rate, at a desired confidence level, for counting experiments where the signal and background levels are small enough to require the use of Poisson statistics. This method also includes the uncertainties on the signal efficiencies and the number of expected background events.

If the mean signal production is s events with an associated background of b events then the probability of observing n events is the the product of the individual Poisson probabilities for the signal and background, summed over all possible combinations satisfying $n_s + n_b = n$,

$$P(n|s+b) = e^{-(s+b)}\frac{(s+b)^n}{n!} = \sum_{n_b=0}^{n}\sum_{n_s=0}^{n-n_b} P(n_b|b)P(n_s|s)\ , \tag{A.1}$$

where $n_{s(b)}$ is the number of signal (background) events observed.

If N events have been observed, the confidence level (CL) for excluding the possibility of simultaneous presence of signal production and background (the $s+b$ hypothesis), is

$$\mathrm{CL}_{s+b} = P(n \le N|s+b) = \sum_{n=0}^{N} P(n|s+b)\ , \tag{A.2}$$

i.e., the probability that the number of events produced, assuming the presence of both signal and background at their hypothesized levels, would be less than, or equal to, that observed in the data.

The confidence level $(1-\mathrm{CL}_{s+b})$ may be used to quote exclusion limits, although it has the disturbing property that if too few candidates are observed to account for the estimated background, then any signal, even the background itself, may be excluded at high confidence. This situation is avoided by noting n_b can only take on values $n_b \le N$, so $P(n_b|b)$, in Equation A.1, must be renormalized to the new range of n_b,

$$P(n_b|b) \mapsto \frac{P(n_b|b)}{\sum_{n_b=0}^{N} P(n_b|b)}\ . \tag{A.3}$$

The dominator can be seen to be CL_b, the probability that the number of events produced by the background alone would be less than, or equal to, that observed in the data. With this change, Equation A.2 becomes,

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{\sum_{n=0}^{N} P(n|s+b)}{\sum_{n_b=0}^{N} P(n_b|b)} = \frac{e^{-(s+b)}\sum_{n=0}^{N}\frac{(s+b)^n}{n!}}{e^{-b}\sum_{n_b=0}^{N}\frac{b^n}{n!}}\ , \tag{A.4}$$

where we have introduced the Poisson probabilities, $P(n|\lambda) = e^{-\lambda}\lambda^n/n!$. Equation A.4 can be used to solve for an the upper limit on the parameter s yielding a confidence level $(1-\mathrm{CL}_s)$.

This result is mathematically identical to the well know results recommended by the Particle Data Group [80], but the interpretations are quite different [115]. This *Modified Frequentist* confidence level is interpreted as follows: for an infinitely large number of experiments, looking for a signal with expectation s and a Poisson distributed background with mean b, where the background is restricted to values of $n_b \le N$, the frequency of observing N or less events is $(1-\mathrm{CL}_s)$.

Writing the number of signal events as $s = \eta\sigma_{\rm BR}\mathcal{L}$, where $\sigma_{\rm BR} = \sigma \times {\rm BR}(A^0 \to b\bar{b})$ is the combined cross-section times branching ratio for the signal, η its detection efficiency and $\mathcal{L}$ the integrated luminosity analysed, then uncertainties in the background b and detection efficiency η can be accounted for [116] by convoluting the terms in Equation A.4,

$$CL_s = \frac{\int_0^\infty \int_0^\infty g(b,b')g(\eta,\eta')\, e^{-(\eta'\sigma_{\rm BR}\mathcal{L}+b')} \sum_{n=0}^{N} \frac{(\eta'\sigma_{\rm BR}\mathcal{L}+b')^n}{n!}\, db'd\eta'}{\int_0^\infty g(b,b')\, e^{-b'} \sum_{n=0}^{N} \frac{(b')^n}{n!}\, db'}, \qquad (A.5)$$

with the resolution functions, here taken to be gaussian,

$$\begin{aligned} g(b,b') &= e^{-(b-b')^2/2\sigma_b^2}/\sigma_b\sqrt{2\pi}, \\ g(\eta,\eta') &= e^{-(\eta-\eta')^2/2\sigma_\eta^2}/\sigma_\eta\sqrt{2\pi}. \end{aligned}$$

σ_b and σ_η are the errors in the number of background events and signal efficiency, which include both systematic and statistical uncertainties.

One can now numerically solve Equation A.5 for $\sigma_{\rm BR}$, which yields the desired CL_s, given the observed values of $\eta \pm \sigma_\eta$ and $b \pm \sigma_b$. This will be interpreted as the maximum cross-section times branching ratio permitted for any signal process present in the data.

Appendix B

The OPAL Collaboration

The OPAL Collaboration

G. Abbiendi[2], K. Ackerstaff[8], G. Alexander[23], J. Allison[16], N. Altekamp[5], K.J. Anderson[9], S. Anderson[12], S. Arcelli[17], S. Asai[24], S.F. Ashby[1], D. Axen[29], G. Azuelos[18,a], A.H. Ball[8], E. Barberio[8], R.J. Barlow[16], J.R. Batley[5], S. Baumann[3], J. Bechtluft[14], T. Behnke[27], K.W. Bell[20], G. Bella[23], A. Bellerive[9], S. Bentvelsen[8], S. Bethke[14], S. Betts[15], O. Biebel[14], A. Biguzzi[5], I.J. Bloodworth[1], P. Bock[11], J. Böhme[14], D. Bonacorsi[2], M. Boutemeur[33], S. Braibant[8], P. Bright-Thomas[1], L. Brigliadori[2], R.M. Brown[20], H.J. Burckhart[8], P. Capiluppi[2], R.K. Carnegie[6], A.A. Carter[13], J.R. Carter[5], C.Y. Chang[17], D.G. Charlton[1,b], D. Chrisman[4], C. Ciocca[2], P.E.L. Clarke[15], E. Clay[15], I. Cohen[23], J.E. Conboy[15], O.C. Cooke[8], J. Couchman[15], C. Couyoumtzelis[13], R.L. Coxe[9], M. Cuffiani[2], S. Dado[22], G.M. Dallavalle[2], R. Davis[30], S. De Jong[12], A. de Roeck[8], P. Dervan[15], K. Desch[27], B. Dienes[32,h], M.S. Dixit[7], J. Dubbert[33], E. Duchovni[26], G. Duckeck[33], I.P. Duerdoth[16], P.G. Estabrooks[6], E. Etzion[23], F. Fabbri[2], A. Fanfani[2], M. Fanti[2], A.A. Faust[30], L. Feld[10], F. Fiedler[27], M. Fierro[2], I. Fleck[10], A. Frey[8], A. Fürtjes[8], D.I. Futyan[16], P. Gagnon[7], J.W. Gary[4], J. Gascon[18], G. Gaycken[27], C. Geich-Gimbel[3], G. Giacomelli[2], P. Giacomelli[2], V. Gibson[5], W.R. Gibson[13], D.M. Gingrich[30,a], D. Glenzinski[9], J. Goldberg[22], W. Gorn[4], C. Grandi[2], K. Graham[28], E. Gross[26], J. Grunhaus[23], M. Gruwé[27], C. Hajdu[31] G.G. Hanson[12], M. Hansroul[8], M. Hapke[13], K. Harder[27], A. Harel[22], C.K. Hargrove[7], M. Harin-Dirac[4], M. Hauschild[8], C.M. Hawkes[1], R. Hawkings[27], R.J. Hemingway[6], G. Herten[10], R.D. Heuer[27], M.D. Hildreth[8], J.C. Hill[5], P.R. Hobson[25], A. Hocker[9], K. Hoffman[8], R.J. Homer[1], A.K. Honma[28,a], D. Horváth[31,c], K.R. Hossain[30], R. Howard[29], P. Hüntemeyer[27], P. Igo-Kemenes[11], D.C. Imrie[25], K. Ishii[24], F.R. Jacob[20], A. Jawahery[17], H. Jeremie[18], M. Jimack[1], C.R. Jones[5], P. Jovanovic[1], T.R. Junk[6], N. Kanaya[24], J. Kanzaki[24], D. Karlen[6], V. Kartvelishvili[16], K. Kawagoe[24], T. Kawamoto[24], P.I. Kayal[30], R.K. Keeler[28], R.G. Kellogg[17], B.W. Kennedy[20], D.H. Kim[19], A. Klier[26], T. Kobayashi[24], M. Kobel[3,d], T.P. Kokott[3], M. Kolrep[10], S. Komamiya[24], R.V. Kowalewski[28], T. Kress[4], P. Krieger[6], J. von Krogh[11], T. Kuhl[3], P. Kyberd[13], G.D. Lafferty[16], H. Landsman[22], D. Lanske[14], J. Lauber[15], I. Lawson[28], J.G. Layter[4], D. Lellouch[26], J. Letts[12], L. Levinson[26], R. Liebisch[11], B. List[8], C. Littlewood[5], A.W. Lloyd[1], S.L. Lloyd[13], F.K. Loebinger[16], G.D. Long[28], M.J. Losty[7], J. Lu[29], J. Ludwig[10], D. Liu[12], A. Macchiolo[18], A. Macpherson[30], W. Mader[3], M. Mannelli[8], S. Marcellini[2], A.J. Martin[13], J.P. Martin[18], G. Martinez[17], T. Mashimo[24], P. Mättig[26], W.J. McDonald[30], J. McKenna[29], E.A. Mckigney[15], T.J. McMahon[1], R.A. McPherson[28], F. Meijers[8], P. Mendez-Lorenzo[33], F.S. Merritt[9], H. Mes[7], A. Michelini[2], S. Mihara[24], G. Mikenberg[26], D.J. Miller[15], W. Mohr[10], A. Montanari[2], T. Mori[24], K. Nagai[8], I. Nakamura[24], H.A. Neal[12,g], R. Nisius[8], S.W. O'Neale[1], F.G. Oakham[7], F. Odorici[2], H.O. Ogren[12], A. Okpara[11], M.J. Oreglia[9], S. Orito[24], G. Pásztor[31], J.R. Pater[16], G.N. Patrick[20], J. Patt[10], R. Perez-Ochoa[8], S. Petzold[27], P. Pfeifenschneider[14], J.E. Pilcher[9], J. Pinfold[30], D.E. Plane[8], P. Poffenberger[28], B. Poli[2], J. Polok[8], M. Przybycień[8,e], A. Quadt[8], C. Rembser[8], H. Rick[8], S. Robertson[28], S.A. Robins[22], N. Rodning[30], J.M. Roney[28], S. Rosati[3], K. Roscoe[16], A.M. Rossi[2], Y. Rozen[22], K. Runge[10], O. Runolfsson[8], D.R. Rust[12], K. Sachs[10], T. Saeki[24], O. Sahr[33], W.M. Sang[25], E.K.G. Sarkisyan[23], C. Sbarra[29], A.D. Schaile[33], O. Schaile[33], P. Scharff-Hansen[8], J. Schieck[11], S. Schmitt[11], A. Schöning[8], M. Schröder[8], M. Schumacher[3], C. Schwick[8], W.G. Scott[20], R. Seuster[14], T.G. Shears[8], B.C. Shen[4],

C.H. Shepherd-Themistocleous[5], P. Sherwood[15], G.P. Siroli[2], A. Sittler[27], A. Skuja[17], A.M. Smith[8], G.A. Snow[17], R. Sobie[28], S. Söldner-Rembold[10,f], S. Spagnolo[20], M. Sproston[20], A. Stahl[3], K. Stephens[16], J. Steuerer[27], K. Stoll[10], D. Strom[19], R. Ströhmer[33], B. Surrow[8], S.D. Talbot[1], P. Taras[18], S. Tarem[22], R. Teuscher[9], M. Thiergen[10], J. Thomas[15], M.A. Thomson[8], E. Torrence[8], S. Towers[6], I. Trigger[18], Z. Trócsányi[32], E. Tsur[23], M.F. Turner-Watson[1], I. Ueda[24], R. Van Kooten[12], P. Vannerem[10], M. Verzocchi[8], H. Voss[3], F. Wäckerle[10], A. Wagner[27], C.P. Ward[5], D.R. Ward[5], P.M. Watkins[1], A.T. Watson[1], N.K. Watson[1], P.S. Wells[8], N. Wermes[3], D. Wetterling[11] J.S. White[6], G.W. Wilson[16], J.A. Wilson[1], T.R. Wyatt[16], S. Yamashita[24], V. Zacek[18], D. Zer-Zion[8]

[1]School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK
[2]Dipartimento di Fisica dell' Università di Bologna and INFN, I-40126 Bologna, Italy
[3]Physikalisches Institut, Universität Bonn, D-53115 Bonn, Germany
[4]Department of Physics, University of California, Riverside CA 92521, USA
[5]Cavendish Laboratory, Cambridge CB3 0HE, UK
[6]Ottawa-Carleton Institute for Physics, Department of Physics, Carleton University, Ottawa, Ontario K1S 5B6, Canada
[7]Centre for Research in Particle Physics, Carleton University, Ottawa, Ontario K1S 5B6, Canada
[8]CERN, European Organisation for Particle Physics, CH-1211 Geneva 23, Switzerland
[9]Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago IL 60637, USA
[10]Fakultät für Physik, Albert Ludwigs Universität, D-79104 Freiburg, Germany
[11]Physikalisches Institut, Universität Heidelberg, D-69120 Heidelberg, Germany
[12]Indiana University, Department of Physics, Swain Hall West 117, Bloomington IN 47405, USA
[13]Queen Mary and Westfield College, University of London, London E1 4NS, UK
[14]Technische Hochschule Aachen, III Physikalisches Institut, Sommerfeldstrasse 26-28, D-52056 Aachen, Germany
[15]University College London, London WC1E 6BT, UK
[16]Department of Physics, Schuster Laboratory, The University, Manchester M13 9PL, UK
[17]Department of Physics, University of Maryland, College Park, MD 20742, USA
[18]Laboratoire de Physique Nucléaire, Université de Montréal, Montréal, Quebec H3C 3J7, Canada
[19]University of Oregon, Department of Physics, Eugene OR 97403, USA
[20]CLRC Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK
[22]Department of Physics, Technion-Israel Institute of Technology, Haifa 32000, Israel
[23]Department of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
[24]International Centre for Elementary Particle Physics and Department of Physics, University of Tokyo, Tokyo 113-0033, and Kobe University, Kobe 657-8501, Japan
[25]Institute of Physical and Environmental Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
[26]Particle Physics Department, Weizmann Institute of Science, Rehovot 76100, Israel
[27]Universität Hamburg/DESY, II Institut für Experimental Physik, Notkestrasse 85, D-22607 Hamburg, Germany
[28]University of Victoria, Department of Physics, P O Box 3055, Victoria BC V8W 3P6, Canada
[29]University of British Columbia, Department of Physics, Vancouver BC V6T 1Z1, Canada
[30]University of Alberta, Department of Physics, Edmonton AB T6G 2J1, Canada
[31]Research Institute for Particle and Nuclear Physics, H-1525 Budapest, P O Box 49, Hungary
[32]Institute of Nuclear Research, H-4001 Debrecen, P O Box 51, Hungary
[33]Ludwigs-Maximilians-Universität München, Sektion Physik, Am Coulombwall 1, D-85748 Garching, Germany
[a] and at TRIUMF, Vancouver, Canada V6T 2A3

[b] and Royal Society University Research Fellow
[c] and Institute of Nuclear Research, Debrecen, Hungary
[d] on leave of absence from the University of Freiburg
[e] and University of Mining and Metallurgy, Cracow
[f] and Heisenberg Fellow
[g] now at Yale University, Dept of Physics, New Haven, USA
[h] and Depart of Experimental Physics, Lajos Kossuth University, Debrecen, Hungary.